Guard Sequence Optimization for Block Transmission over Linear Frequency-Selective Channels

Naofal Al-Dhahir and Suhas N. Diggavi *

Abstract

We show that the optimum length–ν guard sequence for block transmission over a linear Gaussian–noise dispersive channel with memory ν is a *linear combination* of the N information symbols of the block. A closed–form expression for the optimum guard sequence is derived subject to a total average energy constraint on the information and guard symbols. The achievable channel block throughput with the optimum guard sequence is compared with that achievable with two common guard sequence types, namely *zero stuffing* and *cyclic prefix*.

Keywords : Throughput, Transmitters, Intersymbol Interference.

I. Introduction

Maximizing the throughput of block–based transmission over dispersive noisy media is a common challenge for current and future packet–based communication networks. Block–by–block processing has gained increasing popularity recently as an effective receiver architecture for mitigating channel impairments [9, 4, 8, 6, 1]. However, the memory of a dispersive channel causes inter–block interference (IBI) which degrades performance significantly [13]. To eliminate IBI over a channel with memory ν, a guard sequence of length ν is inserted between successive length–N information blocks at the transmitter. Guard sequences most commonly used in practice belong to one of two types. The first type is *zero stuffing* [9, 1] where ν zeros are inserted to clear the channel memory after each block transmission. The second type is *cyclic prefix* [12] where the last ν symbols of every length–N information block are inserted at the beginning of the block (assuming $N > \nu$). Whatever the choice of the guard sequence is, it results in a channel block throughput loss factor of $\frac{\nu}{N+\nu}$ which becomes negligible for $N \gg \nu$. As the input symbol rate increases

*The authors are with AT&T Shannon Laboratory, Florham Park NJ 07932. This paper was presented in part at Globecom'00 in San Francisco.

(wide–band transmission), the channel memory ν increases. Equalization has been proposed as an effective channel shortening technique (see [3] and the references therein), however, it increases the receiver complexity considerably. Moreover, using a large blocklength N to reduce the guard sequence overhead results in increased processing delay (which could be unacceptable for voice and real–time data applications), increased storage requirements, and increased processing power. In addition, on slowly–time–varying channels, blocklengths on the order of the channel coherence time or longer render the common assumption of a time–invariant channel within the block invalid. This, in turn, calls for the use of more complex adaptive processing techniques at the receiver to track channel variations within the block. In summary, for many short–burst transmission scenarios over highly–dispersive channels, the assumption $N \gg \nu$ is not valid.

Recently, it was shown in [2] that given a fixed guard sequence overhead, the channel throughput is maximized by transmitting the information symbols in short and approximately equal–size blocks separated by the guard symbols. This further motivates the investigation of short–block transmission scenarios in more detail.

In this paper, we revisit the choice of the guard sequence for scenarios where $0.1 \leq \frac{\nu}{N} \leq 1.0$ and the channel state information (CSI) is available at the transmitter as in wireline communication systems. Our rationale is simple. The only requirement on the guard sequence is being a length–ν sequence that carries no new information (i.e. it is a *deterministic* function of the N information symbols). While *zero-stuffing* and *cyclic prefix* types of guard sequences result in simple implementations, they unnecessarily constrain the choice of the transmitted input blocks in their $(N+\nu)$–dimensional vector space, hence limiting the achievable channel block throughput. In this paper, we show that part of the channel block throughput loss due to the guard sequence overhead can be recovered by optimizing the choice of the length–ν guard sequence to maximize the channel block throughput.

This paper has two main contributions. First, we prove that, under a channel block throughput maximization criterion, we can restrict the guard sequence to be a linear combination of the information sequence without loss of optimality. Second, we derive a new closed–form expression for the optimum guard sequence when the auto–correlation matrix of the information symbols is fixed.

The rest of this paper is organized as follows. In Section II, we formulate the channel block throughput maximization problem as a function of the guard sequence and the auto–correlation matrix of the information block. In Section III, closed–form solutions are presented under several transmission scenarios. Connections to Vector Coding and Discrete Multitone transmission schemes are discussed in Section IV. Section V presents some numerical results and the paper is concluded in Section VI.

II. Problem formulation

II-A. Input Output Model

We adopt the standard discrete–time representation of an additive–noise linear dispersive channel given by

$$y_k = \sum_{m=0}^{\nu} h_m x_{k-m} + n_k \ , \tag{1}$$

where h_m is the m^{th} coefficient of the channel impulse response that has a memory of ν. The input symbols $\{x_k\}$ and the noise symbols $\{n_k\}$ are assumed to be complex, zero–mean, and have Hermitian positive–definite auto–correlation matrices denoted by $\mathbf{R}_{xx}$ and $\mathbf{R}_{nn}$, respectively. Furthermore, the noise symbols are assumed Gaussian distributed and independent of the input symbols. We consider block transmission scenarios where block–by–block processing is employed at the receiver to detect the information–bearing blocks. Over a block of N output symbols, (1) can be written as follows

$$\begin{bmatrix} y_{k+N-1} \\ y_{k+N-2} \\ \vdots \\ y_k \end{bmatrix} = \begin{bmatrix} h_0 & h_1 & \cdots & h_\nu & 0 & \cdots & 0 \\ 0 & h_0 & h_1 & \cdots & h_\nu & 0 & \cdots \\ \vdots & & & & & & \vdots \\ 0 & \cdots & 0 & h_0 & h_1 & \cdots & h_\nu \end{bmatrix} \begin{bmatrix} x_{k+N-1} \\ x_{k+N-2} \\ \vdots \\ x_{k-\nu} \end{bmatrix} + \begin{bmatrix} n_{k+N-1} \\ n_{k+N-2} \\ \vdots \\ n_k \end{bmatrix} , \tag{2}$$

or more compactly

$$\mathbf{y}_{k+N-1:k} = \mathbf{H}\mathbf{x}_{k+N-1:k-\nu} + \mathbf{n}_{k+N-1:k} \ . \tag{3}$$

To eliminate IBI, thus allowing successive blocks to be processed independently, the first ν symbols in each block, denoted by $\mathbf{x}_{k-1:k-\nu}$, do not carry any new information. They serve as a *guard sequence* that clears the channel memory after each block transmission. It is worth emphasizing that the output symbols corresponding to these guard symbols, namely $\mathbf{y}_{k-1:k-\nu}$, contain interference from the previous block and hence are discarded at the receiver, as illustrated in Figure 1.

In this paper, we consider the most general form for the redundant guard symbols where they are expressed as a *deterministic* function of the information symbols. Therefore, we assume the following mathematical model for the guard sequence

$$\mathbf{x}_{k-1:k-\nu} = f(\mathbf{x}_{k+N-1:k}) \ , \tag{4}$$

where the function $f(.)$ is at first allowed to be non–linear. In the next subsection, we show that $f(.)$ can be assumed linear without loss of optimality, as far as channel block throughput maximization is concerned.

II-B. Channel Block Throughput

The performance measure to be maximized in this paper is the mutual information between the input and output blocks which we shall also refer to as the channel block throughput and denote by $I(\mathbf{X};\mathbf{Y})$, where we have defined $\mathbf{X} \stackrel{def}{=} \mathbf{x}_{k+N-1:k-\nu}$ and $\mathbf{Y} \stackrel{def}{=} \mathbf{y}_{k+N-1:k}$.

Result 1 **:**

When maximizing $I(\mathbf{X};\mathbf{Y})$, we can assume $f(.)$ in (4) to be a linear function without loss of optimality.

Proof : see Appendix A.

For the remainder of this paper, we assume that the guard symbols $\mathbf{x}_{k-1:k-\nu}$ are related to the information symbols $\mathbf{x}_{k+N-1:k}$ by the linear relation

$$\mathbf{x}_{k-1:k-\nu} = \mathbf{G}\mathbf{x}_{k+N-1:k} \ , \tag{5}$$

where $\mathbf{G}$ is a $\nu \times N$ matrix. Multiplying both sides of (3) by the noise–whitening filter $\mathbf{R}_{nn}^{-\frac{1}{2}}$, we get

$$\begin{aligned} \tilde{\mathbf{y}}_{k+N-1:k} &\stackrel{def}{=} \mathbf{R}_{nn}^{-\frac{1}{2}}\mathbf{y}_{k+N-1:k} = \mathbf{R}_{nn}^{-\frac{1}{2}}\mathbf{H}\begin{bmatrix}\mathbf{I}_N \\ \mathbf{G}\end{bmatrix}\mathbf{x}_{k+N-1:k} + \mathbf{R}_{nn}^{-\frac{1}{2}}\mathbf{n}_{k+N-1:k} \\ &= \tilde{\mathbf{H}}\mathbf{x}_{k+N-1:k} + \tilde{\mathbf{n}}_{k+N-1:k} \ , \end{aligned} \tag{6}$$

where $\mathbf{I}_N$ is the identity matrix of size N and $\tilde{\mathbf{n}}_{k+N-1:k} \stackrel{def}{=} \mathbf{R}_{nn}^{-\frac{1}{2}}\mathbf{n}_{k+N-1:k}$ is a unit–energy white–noise sequence (i.e., $\mathbf{R}_{\tilde{n}\tilde{n}} = \mathbf{I}_N$). The matrix $\tilde{\mathbf{H}} \stackrel{def}{=} \mathbf{R}_{nn}^{-\frac{1}{2}}\mathbf{H}\begin{bmatrix}\mathbf{I}_N \\ \mathbf{G}\end{bmatrix}$ is an $N \times N$ white–noise–equivalent channel matrix that depends on the original channel matrix $\mathbf{H}$, the noise auto–correlation matrix $\mathbf{R}_{nn}$, and on the linear mapping matrix $\mathbf{G}$ between the information symbols and the guard symbols.

The block throughput of the white–Gaussian–noise channel in (6) is given by

$$I(\mathbf{X};\mathbf{Y}) \;=\; \frac{N}{N+\nu} I(\mathbf{x}_{k+N-1:k-\nu};\mathbf{y}_{k+N-1:k}) \quad : \ \text{including guard sequence overhead}$$

$$
\begin{aligned}
&= \frac{N}{N+\nu} I(\mathbf{x}_{k+N-1:k}; \mathbf{y}_{k+N-1:k}) \ : \ \text{since guard symbols do not carry new information} \\
&= \frac{N}{N+\nu} I(\mathbf{x}_{k+N-1:k}; \tilde{\mathbf{y}}_{k+N-1:k}) \ : \ \text{since } \mathbf{R}_{nn} \text{ is assumed invertible} \\
&= \frac{N}{N+\nu} \log det(\mathbf{R}_{\tilde{y}\tilde{y}}) \quad : \quad \text{see [5] for a proof ,}
\end{aligned} \tag{7}
$$

where $det(.)$ denotes the determinant of a matrix. The unit of $I(\mathbf{X};\mathbf{Y})$ depends on the logarithm base used. In this paper, we assume base 2 logarithms, hence, $I(\mathbf{X};\mathbf{Y})$ is in bits. Using (6), we can express $\mathbf{R}_{\tilde{y}\tilde{y}}$ as follows

$$
\mathbf{R}_{\tilde{y}\tilde{y}} \stackrel{def}{=} \mathrm{E}[\tilde{\mathbf{y}}_{k+N-1:k}\tilde{\mathbf{y}}^*_{k+N-1:k}] = \tilde{\mathbf{H}}\mathbf{R}_{xx}\tilde{\mathbf{H}}^* + \mathbf{I}_N \ , \tag{8}
$$

where E[.] denotes the expected value and $(.)^*$ denotes the complex–conjugate transpose. Finally, combining (7) with (8), we have

$$
I(\mathbf{X};\mathbf{Y}) = \frac{N}{N+\nu} \log det(\mathbf{I}_N + \mathbf{R}_{xx}^{\frac{*}{2}} \tilde{\mathbf{H}}^* \tilde{\mathbf{H}} \mathbf{R}_{xx}^{\frac{1}{2}}) \tag{9}
$$

$$
= \frac{N}{N+\nu} \log det(\mathbf{I}_N + \mathbf{R}_{xx}^{\frac{*}{2}} \begin{bmatrix} \mathbf{I}_N & \mathbf{G}^* \end{bmatrix} \mathbf{H}^* \mathbf{R}_{nn}^{-1} \mathbf{H} \begin{bmatrix} \mathbf{I}_N \\ \mathbf{G} \end{bmatrix} \mathbf{R}_{xx}^{\frac{1}{2}}) \ . \tag{10}
$$

Our objective is to maximize the channel block throughput given in (10) above by optimizing the ν guard symbols, or equivalently the matrix $\mathbf{G}$, and the auto–correlation matrix of the N information symbols $\mathbf{R}_{xx}$, subject to an average energy constraint on the information and guard symbols.

III. Problem Solution

III-A. Joint $\mathbf{R}_{xx}$ and $\mathbf{G}$ Optimization

We want to solve the constrained joint optimization problem

$$
\max_{\mathbf{G},\mathbf{R}_{xx}} \log det(\mathbf{I}_N + \begin{bmatrix} \mathbf{R}_{xx}^{\frac{*}{2}} & \mathbf{R}_{xx}^{\frac{*}{2}} \mathbf{G}^* \end{bmatrix} \mathbf{H}^* \mathbf{R}_{nn}^{-1} \mathbf{H} \begin{bmatrix} \mathbf{R}_{xx}^{\frac{1}{2}} \\ \mathbf{G}\mathbf{R}_{xx}^{\frac{1}{2}} \end{bmatrix}) \ , \tag{11}
$$

subject to the average energy constraint

$$
\frac{1}{N+\nu} trace(\mathrm{E}[\mathbf{x}_{k+N-1:k-\nu}\mathbf{x}^*_{k+N-1:k-\nu}]) = 1 \tag{12}
$$

$$
\Rightarrow \frac{1}{N+\nu} trace(\mathrm{E} \begin{bmatrix} \mathbf{x}_{k+N-1:k} \\ \mathbf{G}\mathbf{x}_{k+N-1:k} \end{bmatrix} \begin{bmatrix} \mathbf{x}^*_{k+N-1:k} & \mathbf{x}^*_{k+N-1:k}\mathbf{G}^* \end{bmatrix}) = 1
$$

$$\Rightarrow \frac{1}{N+\nu} trace(\begin{bmatrix} \mathbf{R}_{xx} & \mathbf{R}_{xx}\mathbf{G}^* \\ \mathbf{G}\mathbf{R}_{xx} & \mathbf{G}\mathbf{R}_{xx}\mathbf{G}^* \end{bmatrix}) = 1$$

$$\Rightarrow trace(\mathbf{R}_{xx}) + trace(\mathbf{G}\mathbf{R}_{xx}\mathbf{G}^*) = N+\nu \ . \qquad (13)$$

Note that the energy constraint is over the entire transmission block and therefore includes all the well–known choices of guard sequence such as zero stuffing and cyclic prefix as special cases.

Result 2 **:**

Define the eigen–decomposition

$$\mathbf{H}^*\mathbf{R}_{nn}^{-1}\mathbf{H} = \mathbf{U}\begin{bmatrix} \mathbf{\Delta} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{U}^* \ .$$

Then, the jointly–optimum $\mathbf{G}$ and $\mathbf{R}_{xx}$ satisfy

$$\begin{bmatrix} \mathbf{R}_{xx}^{\frac{1}{2}} \\ \mathbf{G}\mathbf{R}_{xx}^{\frac{1}{2}} \end{bmatrix} = \mathbf{U}\begin{bmatrix} \mathbf{\Gamma} \\ \mathbf{0} \end{bmatrix}\mathbf{V}^* \ , \qquad (14)$$

where $\mathbf{V}$ is an arbitrary orthogonal matrix. Hence (14) defines a family of solutions which achieve the same maximum channel block throughput given by

$$I_{opt}(\mathbf{X}\,;\mathbf{Y}) = \frac{N}{N+\nu}\log det(\mathbf{I}_N + \mathbf{\Gamma}^2\mathbf{\Delta})$$

$$= \frac{N}{N+\nu}\sum_{i=1}^{N}\log(1+\gamma_i^2\delta_i) \ . \qquad (15)$$

The N elements of the diagonal matrix $\mathbf{\Gamma}$, denoted by γ_i, are computed using the well–known *water-pour* procedure [5].

Proof : see [1].

Remark :

Note that we deliberately decouple the effects of the two linear matrix filters $\mathbf{R}_{xx}^{\frac{1}{2}}$ and $\mathbf{G}$ because the first acts on the information symbols while the second acts on the guard symbols. This decoupling is in contrast to the approach taken in [1] in which the two filtering effects are lumped together. While the two approaches are equivalent when both $\mathbf{G}$ and $\mathbf{R}_{xx}$ are jointly optimized, decoupling the two effects is critical when only one of the two filters is optimized, as described in the next subsection.

III-B. $\mathbf{G}$ *Optimization for Fixed* $\mathbf{R}_{xx}$

In many practical scenarios, joint optimization of $\mathbf{G}$ and $\mathbf{R}_{xx}$ might not be feasible. One example is the scenario when the information symbols are correlated by passing them through a linear *time-invariant* transmit filter to generate given desirable input spectral characteristics, making $\mathbf{R}_{xx}$ a *fixed* matrix. Another scenario is when the complexity of computing and implementing the jointly–optimum $\mathbf{G}$ and $\mathbf{R}_{xx}$ (in the form of a precoder [1]) is too high. In such scenarios, optimizing $\mathbf{G}$ only offers a good performance–complexity tradeoff and is summarized by the following result

Result 3

Define the matrix partitioning

$$\mathbf{H}^*\mathbf{R}_{nn}^{-1}\mathbf{H} = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{12}^* & \mathbf{R}_{22} \end{bmatrix} , \tag{16}$$

where $\mathbf{R}_{11}$ is the $N \times N$ leading submatrix of $\mathbf{H}^*\mathbf{R}_{nn}^{-1}\mathbf{H}$. Furthermore, define the following matrices

$$\begin{aligned} \mathbf{G}_0 &= \mathbf{R}_{22}^{-1}\mathbf{R}_{12}^*\mathbf{R}_{xx}^{\frac{1}{2}} && (17)\\ \bar{\mathbf{G}} &= \mathbf{G} + \mathbf{G}_0 && (18)\\ \mathbf{F} &= \mathbf{I}_N + \mathbf{R}_{xx}^{\frac{*}{2}}\mathbf{R}^{\perp}\mathbf{R}_{xx}^{\frac{1}{2}} && (19)\\ \mathbf{R}^{\perp} &= \mathbf{R}_{11} - \mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{12}^* && (20)\\ \bar{\mathbf{F}} &= \mathbf{R}_{xx}^{-\frac{*}{2}}\mathbf{F}\mathbf{R}_{xx}^{-\frac{1}{2}} = \mathbf{R}_{xx}^{-1} + \mathbf{R}^{\perp} . && (21) \end{aligned}$$

Then, maximizing the channel block throughput, as defined in (10), for a *fixed* $\mathbf{R}_{xx}$ is achieved by

$$\begin{aligned} \mathbf{G}_{opt} &= \bar{\mathbf{G}} - \mathbf{G}_0 \\ &= \mathbf{U}_{R_{22}} \begin{bmatrix} \boldsymbol{\Sigma}_{\bar{G}} & \mathbf{0}_{\nu\times(N-\nu)} \end{bmatrix} \mathbf{U}_{\bar{F}}^* - \mathbf{G}_0 , \qquad (22) \end{aligned}$$

where $\mathbf{U}_{R_{22}}$ and $\mathbf{U}_{\bar{F}}$ are the modal matrices [1] of the Hermitian matrices $\mathbf{R}_{22}$ and $\bar{\mathbf{F}}$, respectively. In other words

$$\begin{aligned} \mathbf{R}_{22} &= \mathbf{U}_{R_{22}}\boldsymbol{\Sigma}_{R_{22}}\mathbf{U}_{R_{22}}^* \ : \ \text{of size } \nu \times \nu \\ \bar{\mathbf{F}} &= \mathbf{U}_{\bar{F}}\boldsymbol{\Sigma}_{\bar{F}}\mathbf{U}_{\bar{F}}^* \ : \ \text{of size } N \times N. \end{aligned}$$

Moreover, the ν elements of the diagonal matrix $\boldsymbol{\Sigma}_{\bar{G}}$ in (22), denoted by $\sigma_{\bar{G},i}$, are computed as the non–negative real roots of a cubic equation given in (45). Finally, the resulting maximized channel

[1] A modal matrix is a unitary matrix of eigenvectors.

block throughput is given by

$$I(\mathbf{X};\mathbf{Y})_{max} = \frac{N}{N+\nu}(\sum_{i=1}^{N}\log(\lambda_i(\mathbf{F})) + \sum_{i=1}^{\nu}\log(1+\sigma_{R_{22},i}\sigma_{\bar{F},i}^{-1}\sigma_{\bar{G},i}^{2}))\ , \qquad (23)$$

where $\lambda_i(.)$ stands for the i^{th} eigenvalue and $\sigma_{\bar{F},i}$ $(1 \leq i \leq \nu)$ are the ν *smallest* singular values [2] of $\bar{\mathbf{F}}$.

Proof: see Appendix B.

III-C. Special Case : General FIR Channels, White Input, and High SNR

At high SNR, Equation (10) simplifies to (after ignoring the $(\mathbf{I}_N + .)$ term and using the matrix identity $det(\mathbf{I}+\mathbf{AB}) = det(\mathbf{I}+\mathbf{BA})$)

$$\begin{aligned} I(\mathbf{X};\mathbf{Y})_{high-SNR} &\approx \frac{N}{N+\nu}\log det(\mathbf{R}_{xx}\begin{bmatrix}\mathbf{I}_N & \mathbf{G}^*\end{bmatrix}\mathbf{H}^*\mathbf{R}_{nn}^{-1}\mathbf{H}\begin{bmatrix}\mathbf{I}_N \\ \mathbf{G}\end{bmatrix}) \\ &= \frac{N}{N+\nu}(\log det(\mathbf{R}_{xx}) + \log det(\begin{bmatrix}\mathbf{I}_N & \mathbf{G}^*\end{bmatrix}\mathbf{H}^*\mathbf{R}_{nn}^{-1}\mathbf{H}\begin{bmatrix}\mathbf{I}_N \\ \mathbf{G}\end{bmatrix}))\ . \qquad (24) \end{aligned}$$

It is clear from (24) that at high SNR, the second term dominates the first term. Hence, the effect of optimizing $\mathbf{R}_{xx}$ on $I(\mathbf{X};\mathbf{Y})$ becomes negligible. For analytical convenience, we assume white input, i.e. we set $\mathbf{R}_{xx} = E_x\mathbf{I}_N$, where E_x is the input energy per dimension.

Therefore, for the white–input case at high SNR, maximizing $I(\mathbf{X};\mathbf{Y})$, as defined in (9), becomes equivalent to[3]

$$\max_{\mathbf{G}}\log|det(\tilde{\mathbf{H}})|^2 = \max_{\mathbf{G}}\log|det(\mathbf{R}_{nn}^{-\frac{1}{2}}\mathbf{H}\begin{bmatrix}\mathbf{I}_N \\ \mathbf{G}\end{bmatrix})|^2\ . \qquad (25)$$

Define $\hat{\mathbf{H}} = \mathbf{R}_{nn}^{-\frac{1}{2}}\mathbf{H} \stackrel{def}{=} \begin{bmatrix}\mathbf{H}_1 & \mathbf{H}_2\end{bmatrix}$, where $\mathbf{H}_1$ consists of the first N columns of $\hat{\mathbf{H}}$ and $\mathbf{H}_2$ consists of its last ν columns. Then, the optimization problem in (25) can be recast as follows [4]

$$\max_{\mathbf{G}}\log|det(\mathbf{I}_N + \mathbf{H}_1^{-1}\mathbf{H}_2\mathbf{G})|^2 \stackrel{def}{=} \max_{\mathbf{G}}\log|det(\mathbf{I}_N + \bar{\mathbf{H}}\mathbf{G})|^2 = \max_{\mathbf{G}}\log|det(\mathbf{I}_\nu + \mathbf{G}\bar{\mathbf{H}})|^2\ , \qquad (26)$$

subject to the average energy constraint in (13). For $\mathbf{R}_{xx} = E_x\mathbf{I}_N$, this constraint simplifies to

$$\begin{aligned} E_xN + E_x trace(\mathbf{G}\mathbf{G}^*) &= N+\nu \\ \Rightarrow trace(\mathbf{G}\mathbf{G}^*) &= \frac{N+\nu}{E_x} - N \geq 0\ . \qquad (27) \end{aligned}$$

[2] $\bar{\mathbf{F}}$ and $\mathbf{R}_{22}$ are Hermitian matrices, hence, their singular values are equal to their eigenvalues.

[3] For a complex matrix $\tilde{\mathbf{H}}$, $det(\tilde{\mathbf{H}}^*) = (det(\tilde{\mathbf{H}}))^*$.

[4] $\mathbf{H}_1$ is a square full–rank matrix, hence, $\mathbf{H}_1^{-1}$ exists.

Note that this analysis includes zero–stuffing as a special case with $\mathbf{G} = \mathbf{0}_{\nu \times N}$ and $E_x = \frac{N+\nu}{N}$.

Result 4

Define the singular value decomposition of the $N \times \nu$ matrix $\bar{\mathbf{H}}$

$$\bar{\mathbf{H}} \stackrel{def}{=} \mathbf{H}_1^{-1}\mathbf{H}_2 = \mathbf{U}_{\bar{H}} \begin{bmatrix} \mathbf{\Sigma}_{\bar{H}} \\ \mathbf{0}_{(N-\nu)\times\nu} \end{bmatrix} \mathbf{V}_{\bar{H}}^* = \mathbf{U}_{\bar{H}}^{(1)} \mathbf{\Sigma}_{\bar{H}} \mathbf{V}_{\bar{H}}^* \;, \tag{28}$$

where $\mathbf{V}_{\bar{H}}$ is a $\nu \times \nu$ unitary matrix and $\mathbf{U}_{\bar{H}}^{(1)}$ consists of the first ν columns of the $N \times N$ unitary matrix $\mathbf{U}_{\bar{H}}$. Similarly, define the singular value decomposition of the $\nu \times N$ matrix $\mathbf{G}$

$$\mathbf{G} = \mathbf{U}_G \begin{bmatrix} \mathbf{\Sigma}_G & \mathbf{0}_{\nu\times(N-\nu)} \end{bmatrix} \mathbf{V}_G^* = \mathbf{U}_G \mathbf{\Sigma}_G \mathbf{V}_G^{*(1)} \;, \tag{29}$$

where $\mathbf{U}_G$ is a $\nu \times \nu$ unitary matrix and $\mathbf{V}_G^{(1)}$ consists of the first ν columns of the $N \times N$ unitary matrix $\mathbf{V}_G$. Then, the optimum $\mathbf{G}$ that solves the channel block throughput maximization problem in (26) satisfies the relations

$$\mathbf{U}_G = \mathbf{V}_{\bar{H}} \;\text{ and }\; \mathbf{V}_G^{(1)} = \mathbf{U}_{\bar{H}}^{(1)} \;. \tag{30}$$

Entries of the $\nu \times \nu$ diagonal matrix $\mathbf{\Sigma}_G$ are computed as the ν non–negative real roots [5] of the ν quadratic equations

$$\alpha \sigma_{\bar{H},i} \sigma_{G,i}^2 + \alpha \sigma_{G,i} + \sigma_{\bar{H},i} = 0 \;\; : \;\; 1 \le i \le \nu \;, \tag{31}$$

where $\sigma_{\bar{H},i}$ are the entries of the $\nu \times \nu$ diagonal matrix $\mathbf{\Sigma}_{\bar{H}}$ and α is a constant determined from the average energy constraint in (27) which is equivalent to the constraint $\sum_{i=1}^{\nu} \sigma_{G,i}^2 = \frac{N+\nu}{E_x} - N$.

Proof : see Appendix C.

Substituting (28) and (29) in (26), we get

$$\begin{aligned} \max_G \log|det(\mathbf{I}_\nu + \mathbf{G}\bar{\mathbf{H}})|^2 &= \max_G \log|det(\mathbf{I}_\nu + \mathbf{U}_G \mathbf{\Sigma}_G \mathbf{V}_G^{*(1)} \mathbf{U}_{\bar{H}}^{(1)} \mathbf{\Sigma}_{\bar{H}} \mathbf{V}_{\bar{H}}^*)|^2 \qquad (32) \\ &= \max_G \log|det(\mathbf{I}_\nu + \mathbf{\Sigma}_{\bar{H}}^{\frac{*}{2}} \mathbf{V}_{\bar{H}}^* \mathbf{U}_G \mathbf{\Sigma}_G \mathbf{V}_G^{*(1)} \mathbf{U}_{\bar{H}}^{(1)} \mathbf{\Sigma}_{\bar{H}}^{\frac{1}{2}})|^2 \\ &= \max_G \sum_{i=1}^{\nu} \log|1 + \lambda_i(\mathbf{P})|^2 \;, \qquad (33) \end{aligned}$$

where we defined $\mathbf{P} \stackrel{def}{=} \mathbf{\Sigma}_{\bar{H}}^{\frac{*}{2}} \mathbf{V}_{\bar{H}}^* \mathbf{U}_G \mathbf{\Sigma}_G \mathbf{V}_G^{*(1)} \mathbf{U}_{\bar{H}}^{(1)} \mathbf{\Sigma}_{\bar{H}}^{\frac{1}{2}}$.

The $\nu \times \nu$ matrix $\mathbf{P}$ is generally non–Hermitian, and hence its eigenvalues are complex in general. However, the optimum choice $\mathbf{U}_G = \mathbf{V}_{\bar{H}}$ and $\mathbf{V}_G^{(1)} = \mathbf{U}_{\bar{H}}^{(1)}$ in (30) makes $\mathbf{P}$ equal to the diagonal matrix $\mathbf{\Sigma}_G \mathbf{\Sigma}_{\bar{H}}$ whose ν diagonal entries (which are also equal to its eigenvalues) are positive real

[5] Since the singular values of a matrix are non–negative real numbers by definition.

numbers. This is intuitively appealing since it is clear that Equation (33) is maximized when the eigenvalues of $\mathbf{P}$ are real valued.

Substituting (30) in (32) we arrive at the following equivalent but simplified optimization problem

$$\max_{G} \log |det(\mathbf{I}_\nu + \mathbf{G}\bar{\mathbf{H}})|^2 = \max_{G} \log |det(\mathbf{I}_\nu + \mathbf{\Sigma}_G \mathbf{\Sigma}_{\bar{H}})|^2 = \max_{\sigma_{G,i}} \sum_{i=1}^{\nu} \log |1 + \sigma_{G,i}\sigma_{\bar{H},i}|^2 \ , \tag{34}$$

subject to the average energy constraint $\sum_{i=1}^{\nu} \sigma_{G,i}^2 = \frac{N+\nu}{E_x} - N$. Using standard Lagrange multiplier techniques, it can be easily shown that the optimum $\sigma_{G,i}$ are computed from (31).

III-D. Special Case : Two-Tap Channels, White Input, and High SNR

Another interesting special case is the two–tap dispersive channel with white input and noise at high SNR with arbitrary blocklength N. Under these assumptions, the optimization problem in (26) simplifies to

$$\max_{\mathbf{G}} \log |1 + \mathbf{G}\bar{\mathbf{H}}|^2 \ \text{ subject to } \ \mathbf{G}\mathbf{G}^* = \frac{N+1}{E_x} - N \ , \tag{35}$$

where $\mathbf{G}$ is now a $1 \times N$ row vector and $\bar{\mathbf{H}}$ is an $N \times 1$ column vector. Using the *Cauchy-Schwartz Inequality*, the optimum $\mathbf{G}$ is given by

$$\mathbf{G}_{opt} = \sqrt{(\frac{N+1}{E_x} - N)} \frac{\bar{\mathbf{H}}^*}{\| \bar{\mathbf{H}} \|} \ : \ \text{assuming white input, high SNR, and } \nu = 1 \ , \tag{36}$$

where $\| \ . \ \|$ denotes the Euclidean norm of a vector. The vector $\bar{\mathbf{H}} = \mathbf{H}_1^{-1}\mathbf{H}_2$ can be expressed directly in terms of the two channel impulse response coefficients h_0 and h_1 by using the upper–triangular Toeplitz structure of the $N \times N$ matrix $\mathbf{H}_1$ and the special form of $\mathbf{H}_2$ in this case, namely, $\mathbf{H}_2 = \begin{bmatrix} 0 & \cdots & 0 & h_1 \end{bmatrix}^t$ (where $(.)^t$ denotes the transpose). More specifically, it is easy to show that the i^{th} element of $\bar{\mathbf{H}}$ which we denote by $\bar{\mathbf{H}}(i)$ is given by

$$\bar{\mathbf{H}}(i) = (-1)^{N-i} (\frac{h_1}{h_0})^{N-i+1} \ : \ i = 1, 2, \cdots, N \ . \tag{37}$$

We define $\mathbf{R}_{nn} = E_n \mathbf{I}_N = \frac{1}{SNR}\mathbf{I}_N$ for the white input and noise case [6]. Hence, a closed–form expression for the maximum achievable channel block throughput in this case with $\mathbf{G}$ optimization

[6] This definition arises from our input energy constraint in (13) which constrains the average energy (per dimension) of the information and guard symbols in each block to be equal to unity.

is given by

$$
\begin{aligned}
I(\mathbf{X};\mathbf{Y})^{2-tap}_{\text{high-SNR}} &= \frac{N}{N+\nu}\log|det(E_x\sqrt{SNR}\mathbf{H}_1)det(\mathbf{I}_N+\bar{\mathbf{H}}\mathbf{G})|^2 \\
&= \frac{N}{N+\nu}(\log|E_x\sqrt{SNR}h_0|^{2N}+\log|1+\mathbf{G}\bar{\mathbf{H}}|^2) \quad : \text{since } det(\mathbf{H}_1)=h_0^N \\
&= \frac{N}{N+\nu}(N\log(SNRE_x^2|h_0|^2)+2\log(1+\sqrt{(\frac{N+1}{E_x}-N)}\parallel\bar{\mathbf{H}}\parallel)) \quad : \text{using (36)}\ .
\end{aligned}
$$

If we define $a \stackrel{def}{=} |\frac{h_0}{h_1}| > 0$, then it is easy to show using Equation (37) that

$$
\parallel \bar{\mathbf{H}} \parallel = \begin{cases} a^{-N}\sqrt{\frac{a^{2N}-1}{a^2-1}} & \text{if } a \neq 1 \\ \sqrt{N} & \text{if } a = 1 \end{cases}
$$

IV. Related Issues

IV-A. Connection with Vector Coding and Discrete Multitone

For white noise, Vector Coding (VC) [9] follows trivially as a special case [7] of the general framework in (6) by setting the guard sequence to be the all–zeros sequence, i.e., $\mathbf{G} = \mathbf{0}_{\nu\times N}$.

In Discrete Multitone (DMT)[12], the length–ν guard sequence is equal to the last ν symbols of the length–N information block, hence the name cyclic prefix. Therefore, DMT follows as a special case of the general framework of this paper by setting $\mathbf{G} = \left[\begin{array}{cc}\mathbf{I}_\nu & \mathbf{0}_{\nu\times(N-\nu)}\end{array}\right]$ in (6). This particular choice of guard sequence is by far the most popular in practice nowadays for two main reasons. First, the cyclic prefix is independent of the channel characteristics which makes it attractive for scenarios where channel knowledge is not available at the transmitter. Second, and more importantly, the cyclic prefix guard sequence makes the equivalent channel matrix $\tilde{\mathbf{H}}$ circulant (assuming the noise correlation matrix to be circulant also) and hence diagonalizable by the computationally–efficient Fast Fourier Transform (FFT). Therefore, for scenarios where $\nu \ll N$, DMT is clearly the preferred choice from an implementation point of view and results in negligible channel block throughput loss from the optimum scheme, as will be shown in the simulation results of Section V. For both VC and DMT, only $\mathbf{R}_{xx}$ needs to be optimized. This optimization is carried out using the water–pour procedure [5].

[7] Note that only when the guard sequence is the all–zeros sequence as in vector coding, the receiver need not discard the last ν samples of the length–$(N+\nu)$ received block since no IBI is present in them. While it can be shown that processing these additional ν samples does not improve the channel block throughput for white noise, the same result does not hold in general for colored noise.

IV-B. Optimized Short Guard Sequences

To reduce throughput loss due to guard sequence overhead, it might be desirable in some transmission scenarios to use a short guard sequence, i.e., one whose length N_b is less than the channel memory ν. If we define the following column partitioning of the channel matrix $\mathbf{H} = \left[\underbrace{\mathbf{H}_1}_{N} \; \underbrace{\mathbf{H}_2}_{N_b} \; \underbrace{\mathbf{H}_3}_{(\nu-N_b)} \right]$ and the submatrix $\bar{\mathbf{H}} = \left[\mathbf{H}_1 \quad \mathbf{H}_2 \right]$, then it can be shown that the channel block throughput optimization problem in (11) becomes

$$\max_{\mathbf{G},\mathbf{R}_{xx}} \log det(\mathbf{I}_N + \mathbf{R}_{xx}^{\frac{*}{2}} \left[\mathbf{I}_N \quad \mathbf{G}^* \right] \bar{\mathbf{H}}^* \mathbf{R}_{zz}^{-1} \bar{\mathbf{H}} \begin{bmatrix} \mathbf{I}_N \\ \mathbf{G} \end{bmatrix} \mathbf{R}_{xx}^{\frac{1}{2}}) \ .$$

The IBI–and–noise auto–correlation matrix $\mathbf{R}_{zz}$ is given by

$$\mathbf{R}_{zz} = \bar{\mathbf{H}}_3 \mathbf{R}_{xx} \bar{\mathbf{H}}_3^* + \mathbf{R}_{nn} \ ,$$

where $\bar{\mathbf{H}}_3 = \mathbf{H}_3 \left[\mathbf{I}_{\nu-N_b} \quad \mathbf{0}_{(\nu-N_b)\times(N-\nu+N_b)} \right]$. Hence, we can apply our previous results simply by replacing $\mathbf{R}_{nn}$ and $\mathbf{H}$ by $\mathbf{R}_{zz}$ and $\bar{\mathbf{H}}$, respectively.

IV-C. Receiver Performance

At the receiver end, the information block $\mathbf{x}_{k+N-1:k}$ is estimated from the output block $\mathbf{y}_{k+N-1:k}$ which is corrupted by channel dispersion and noise. A popular receiver structure is the linear minimum–mean-square–error (MMSE) detector where a *soft* estimate of the input block is formed by multiplying the output block by an $N \times N$ equalizer matrix $\mathbf{W}$ chosen such that the average energy of the error vector between the input block and its estimate is minimized. Using basic principles of linear estimation theory, the optimum equalizer matrix $\mathbf{W}_{opt}$ and the corresponding optimum error auto–correlation matrix are given by [10]

$$\mathbf{W}_{opt} = \mathbf{R}_{\tilde{y}\tilde{y}}^{-1} \mathbf{R}_{\tilde{y}x} \tag{38}$$

$$\mathbf{R}_{ee,min} = \mathbf{R}_{xx} - \mathbf{R}_{\tilde{y}x}^* \mathbf{R}_{\tilde{y}\tilde{y}}^{-1} \mathbf{R}_{\tilde{y}x} = (\mathbf{R}_{xx}^{-1} + \tilde{\mathbf{H}}^* \tilde{\mathbf{H}})^{-1} \ , \tag{39}$$

where $\mathbf{R}_{\tilde{y}x} = \tilde{\mathbf{H}} \mathbf{R}_{xx}$ and the last line follows from Equations (6), (8), and the matrix inversion lemma [10].

Result 5

Define the receiver SNR as $SNR_{rcvr} = det(\mathbf{R}_{ee,min}^{-1} \mathbf{R}_{xx})^{\frac{1}{\bar{N}}}$ where $\bar{N} \leq N$ is the effective number of input dimensions used for transmission and is equal to the number of non–zero eigenvalues in the optimized $\mathbf{R}_{xx}$. Then, the normalized channel block throughput (in bits/symbol) is given by

$$\bar{I}(\mathbf{X};\mathbf{Y}) \stackrel{def}{=} \frac{I(\mathbf{X};\mathbf{Y})}{N} = \frac{\bar{N}}{N+\nu} \log SNR_{rcvr} \ . \tag{40}$$

Moreover, $SNR_{rcvr} = \prod_{i=1}^{\bar{N}}(1+\sigma^2_{\tilde{H},i}\sigma^2_{x,i})^{\frac{1}{N}}$ where $\sigma_{\tilde{H},i}$ and $\sigma_{x,i}$ are the singular values of $\tilde{\mathbf{H}}$ and $\mathbf{R}_{xx}$, respectively.

Proof: See Appendix D.

V. Numerical Results

We start by considering the dispersive channel $h(D) = 0.407 + 0.815D + 0.407D^2$ (also studied in [11]) and normalize its energy to 1. We assume $N = 8$, hence, the guard sequence overhead is $\frac{\nu}{N+\nu} = 20\%$. Furthermore, the noise is assumed white Gaussian; hence, $\mathbf{R}_{nn} = \frac{1}{SNR}\mathbf{I}_N$ where SNR denotes the input SNR level. First, we compare the optimum transmission scheme of Section III–A with the white–input scheme (i.e. $\mathbf{R}_{xx} = E_x\mathbf{I}_N$) at different input SNR levels. The performance measure adopted is the *normalized percentage throughput gain* (NPTG) defined as

$$NPTG = \frac{I_{opt}(\mathbf{X};\mathbf{Y}) - I_{xx}(\mathbf{X};\mathbf{Y})}{I_{xx}(\mathbf{X};\mathbf{Y})} \times 100 \ , \tag{41}$$

where $I_{xx}(\mathbf{X};\mathbf{Y})$ denotes the channel block throughput of the transmission scheme xx under consideration (i.e. xx could be white–input, DMT, or VC in this paper). Figure 2 shows that the performance of the white–input scheme approaches that of the optimum scheme at moderate–to–high input SNR levels, as expected. The performance advantage of the optimum scheme of Section III-A over DMT and VC transmission schemes (with their $\mathbf{R}_{xx}$ optimized) is depicted in Figure 3 as a function of the input SNR level. It is clear that optimization of $\mathbf{G}$ improves channel block throughput at all input SNR levels. Since DMT and VC are special cases of the optimum scheme, they are always inferior to it in throughput. However, the throughput gap between the three schemes diminishes as the blocklength increases (at a fixed ν), as shown in Figure 4.

For the remainder of this section, we investigate the effect of the channel characteristics on the performance. We generate histograms of NPTG for 1000 random channel realizations. The channel impulse responses are generated as FIR filters with $(\nu+1)$ uncorrelated taps where each tap is a zero–mean unit–variance complex Gaussian random variable. For each realization, the channel impulse response energy is normalized to unity. In Figures 5–7, we plot NPTG of optimum scheme of Section III–A over DMT for $N = 8$ and an input SNR of 15 dB. The channel memory takes the values 1, 2, and 3, corresponding to a guard sequence overhead of 11.1%, 20%, and 27.3%, respectively. It is clear from the figures that the optimum scheme of Section III–A always outperforms DMT and the performance advantage increases as ν increases (at a fixed N). Similar trends were observed by comparing the optimum scheme with VC and are omitted here for brevity. Finally, we show in Figure 8 the throughput improvement of the suboptimum scheme of Section III-B over DMT. The $\mathbf{R}_{xx}$ used in both schemes is the optimum $\mathbf{R}_{xx}$ of DMT. Hence, the throughput gain observed in Figure 8 is due to $\mathbf{G}$ optimization. Comparing Figures 6 and 8, we find that the suboptimum scheme achieves an appreciable portion of the performance gain of the optimum scheme. Table I summarizes the effect of N, ν, and input SNR on the choice of $\mathbf{R}_{xx}$ and $\mathbf{G}$ that maximizes the channel block throughput.

VI. Conclusions

We showed that the block throughput of Gaussian–noise dispersive channels can be improved by choosing the guard symbols to be an optimized linear combination of the information symbols within each block. The throughput improvement becomes more appreciable for short–block transmission over channels with long memory. Vector coding and discrete multitone transmission schemes were shown to be special cases of the proposed optimum transmission scheme.

Acknowledgment

We would like to thank Professor G.B. Giannakis from the University of Minnesota for clarifying the connection of this work to [1] and Professor Roger A. Horn from the University of Utah for pointing out the singular–value inequalities in [7] used in the proofs of Results 3 and 4.

Appendix

A. Proof of Result 1

To maximize $I(\mathbf{X};\mathbf{Y})$ for the Gaussian–noise model in (3), the input vector $\mathbf{X} = \mathbf{x}_{k+N-1:k}$ must be Gaussian distributed [5] regardless of any further constraints on its components. In particular, when the guard sequence is assumed to have the functional dependence $\mathbf{x}_{k-1:k-\nu} = f(\mathbf{x}_{k+N-1:k})$, then it must also be Gaussian distributed for $I(\mathbf{X};\mathbf{Y})$ to be maximized. Now, assume that $f(.)$ is nonlinear. Recall that any Gaussian random vector is completely determined by its mean and auto–correlation matrix. Therefore, $\mathbf{x}_{k-1:k-\nu}$ is statistically equivalent to $\tilde{\mathbf{x}}_{k-1:k-\nu} \stackrel{def}{=} f_{linear}(\mathbf{x}_{k+N-1:k})$

where $f_{linear}(.)$ is a linear function chosen such that the Gaussian random vectors $\mathbf{x}_{k-1:k-\nu}$ and $\tilde{\mathbf{x}}_{k-1:k-\nu}$ have the same mean and auto–correlation matrix. □

B. Proof of Result 3

Substituting (16) in (11), we get

$$
\begin{aligned}
& \log det(\mathbf{I}_N + \mathbf{R}_{xx}^{\frac{*}{2}} \begin{bmatrix} \mathbf{I}_N & \mathbf{G}^* \end{bmatrix} \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{12}^* & \mathbf{R}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I}_N \\ \mathbf{G} \end{bmatrix} \mathbf{R}_{xx}^{\frac{1}{2}}) \\
= \; & \log det(\mathbf{F} + \mathbf{R}_{xx}^{\frac{*}{2}} \bar{\mathbf{G}}^* \mathbf{R}_{22} \bar{\mathbf{G}} \mathbf{R}_{xx}^{\frac{1}{2}}) \;:\; \text{where } \bar{\mathbf{G}} \text{ and } \mathbf{F} \text{ are defined in (18) and (19), respectively} \\
= \; & \log det(\mathbf{F}) + \log det(\mathbf{I}_N + \mathbf{F}^{-\frac{*}{2}} \mathbf{R}_{xx}^{\frac{*}{2}} \bar{\mathbf{G}}^* \mathbf{R}_{22} \bar{\mathbf{G}} \mathbf{R}_{xx}^{\frac{1}{2}} \mathbf{F}^{-\frac{1}{2}}) \\
= \; & \log det(\mathbf{F}) + \log det(\mathbf{I}_\nu + \mathbf{R}_{22}^{\frac{1}{2}} \bar{\mathbf{G}} \mathbf{R}_{xx}^{\frac{1}{2}} \mathbf{F}^{-1} \mathbf{R}_{xx}^{\frac{*}{2}} \bar{\mathbf{G}}^* \mathbf{R}_{22}^{\frac{*}{2}}) \\
= \; & \log det(\mathbf{F}) + \log det(\mathbf{I}_\nu + \mathbf{R}_{22}^{\frac{1}{2}} \bar{\mathbf{G}} \bar{\mathbf{F}}^{-1} \bar{\mathbf{G}}^* \mathbf{R}_{22}^{\frac{*}{2}}) \;:\; \text{where } \bar{\mathbf{F}} \text{ is defined in (21)} \,. \qquad (42)
\end{aligned}
$$

Since $\mathbf{F}$ is a constant matrix, maximizing $I(\mathbf{X};\mathbf{Y})$ is equivalent to maximizing the second term in (42) which can be carried out as follows

$$
\begin{aligned}
\log det(\mathbf{I}_\nu + \mathbf{R}_{22}^{\frac{1}{2}} \bar{\mathbf{G}} \bar{\mathbf{F}}^{-1} \bar{\mathbf{G}}^* \mathbf{R}_{22}^{\frac{*}{2}}) \; &= \; \sum_{i=1}^{\nu} \log(1 + \sigma_i^2(\mathbf{R}_{22}^{\frac{1}{2}} \bar{\mathbf{G}} \bar{\mathbf{F}}^{-\frac{1}{2}})) \\
&\leq \; \sum_{i=1}^{\nu} \log(1 + \sigma_i^2(\mathbf{R}_{22}^{\frac{1}{2}}) \sigma_i^2(\bar{\mathbf{G}}) \sigma_i^2(\bar{\mathbf{F}}^{-\frac{1}{2}})) \\
&\stackrel{def}{=} \; \sum_{i=1}^{\nu} \log(1 + \sigma_{R_{22},i} \sigma_{\bar{F},i}^{-1} \sigma_{\bar{G},i}^2) \,, \qquad (43)
\end{aligned}
$$

where the second–to–last line above follows by repeated application of Theorem (3.3.14) part (c) in [7] with $f(x) = \log(1+x^2)$ and $\sigma_i(.)$ stands for the i^{th} singular value. Define the singular value decomposition

$$\bar{\mathbf{G}} = \mathbf{U}_{\bar{G}} \begin{bmatrix} \mathbf{\Sigma}_{\bar{G}} & \mathbf{0}_{\nu \times (N-\nu)} \end{bmatrix} \mathbf{V}_{\bar{G}}^* \,.$$

Then, it is straightforward to show that the maximum in (43) is achieved by setting

$$\mathbf{U}_{\bar{G}} = \mathbf{U}_{R_{22}} \quad ; \quad \mathbf{V}_{\bar{G}} = \mathbf{U}_{\bar{F}} \,,$$

which is the same as (22). Finally, to compute $\sigma_{\bar{G},i}$, the trace constraint in (13) is combined with (18) resulting in the following *quadratic-equation* constraint on $\sigma_{\bar{G},i}$

$$\sum_{i=1}^{\nu} (a_i \sigma_{\bar{\mathbf{G}},i}^2 + b_i \sigma_{\bar{\mathbf{G}},i} + c_i) = (N+\nu) - trace(\mathbf{R}_{xx}) \,, \qquad (44)$$

where a_i, b_i, and c_i are constants that depend on the matrices $\mathbf{H}$, $\mathbf{R}_{xx}$, and $\mathbf{R}_{nn}$. Then, it is straightforward to show (using Lagrange multiplier technique) that maximizing (43) with respect to $\sigma_{\bar{G},i}$ subject to the quadratic–equation constraint in (44) results in the following *cubic* equation in $\sigma_{\bar{G},i}$

$$2a_ik_i\sigma_{\bar{G},i}^3 + b_ik_i\sigma_{\bar{G},i}^2 + 2(a_i + \frac{k_i}{\lambda})\sigma_{\bar{G},i} + b_i = 0 \ , \tag{45}$$

where $k_i \stackrel{def}{=} \sigma_{\mathbf{R}_{22},i}\sigma_{\bar{\mathbf{F}},i}^{-1}$ and λ is a Lagrange constant computed from the trace constraint in (44). □

C. Proof of Result 4

Starting with the right handside of (26), we have the following inequalities

$$\begin{aligned}
\log|det(\mathbf{I}_\nu + \mathbf{G}\bar{\mathbf{H}})|^2 &= \log\prod_{i=1}^{\nu}|1+\lambda_i(\mathbf{G}\bar{\mathbf{H}})|^2 \\
&\le 2\log\prod_{i=1}^{\nu}(1+|\lambda_i(\mathbf{G}\bar{\mathbf{H}})|) \\
&\le 2\log\prod_{i=1}^{\nu}(1+\sigma_i(\mathbf{G}\bar{\mathbf{H}})) : \quad \text{using (3.3.23) of [7]} \\
&\le 2\log\prod_{i=1}^{\nu}(1+\sigma_i(\mathbf{G})\sigma_i(\bar{\mathbf{H}})) \\
&\quad \text{(using Theorem (3.3.14) part (c) in [7] with } f(x)=\log(1+x)) \\
&\stackrel{def}{=} \sum_{i=1}^{\nu}\log(1+\sigma_{G,i}\sigma_{\bar{H},i})^2 \ .□
\end{aligned}$$

D. Proof of Result 5

Starting from Equation (9), we have

$$\bar{I}(\mathbf{X};\mathbf{Y}) \stackrel{def}{=} \frac{I(\mathbf{X};\mathbf{Y})}{N} = \frac{1}{N+\nu}\log det((\mathbf{R}_{xx}^{-1} + \tilde{\mathbf{H}}^*\tilde{\mathbf{H}})\mathbf{R}_{xx}) : \text{in bits/symbol} \tag{46}$$

$$= \frac{\bar{N}}{N+\nu}\log det(\mathbf{R}_{ee,min}^{-1}\mathbf{R}_{xx})^{\frac{1}{N}} : \text{using Equation (39)} \tag{47}$$

$$\stackrel{def}{=} \frac{\bar{N}}{N+\nu}\log SNR_{rcvr} \ , \tag{48}$$

where $SNR_{rcvr} \stackrel{def}{=} det(\mathbf{R}_{ee,min}^{-1}\mathbf{R}_{xx})^{\frac{1}{N}}$ is the *geometric average* of the SNR's in estimating $\mathbf{x}_{k+N-1:k}$ from $\mathbf{y}_{k+N-1:k}$. We can see from (48) that maximizing $I(\mathbf{X};\mathbf{Y})$ is equivalent to maximizing SNR_{rcvr}

at the receiver. Now, when $\mathbf{R}_{xx}$ and $\mathbf{G}$ are optimized to maximize $I(\mathbf{X};\mathbf{Y})$ as shown in Subsection III–A, we get

$$\begin{aligned}\bar{I}(\mathbf{X};\mathbf{Y})_{opt} &= \frac{\bar{N}}{N+\nu}\log(\mathbf{I}_N+\mathbf{\Sigma}_{\tilde{H}}^2\mathbf{\Sigma}_x^2)^{\frac{1}{N}} \\ &= \frac{\bar{N}}{N+\nu}\log(\prod_{i=1}^{\bar{N}}(1+\sigma_{\tilde{H},i}^2\sigma_{x,i}^2))^{\frac{1}{N}}\ . \end{aligned} \tag{49}$$

Comparing (49) with (47), we can relate the receiver performance measure SNR_{rcvr} to the eigenvalues of the equivalent channel matrix and the input auto–correlation matrix as follows

$$SNR_{rcvr} = (\prod_{i=1}^{\bar{N}}(1+\sigma_{\tilde{H},i}^2\sigma_{x,i}^2))^{\frac{1}{N}}\ . \qquad \Box \tag{50}$$

References

[1] A. Scaglione and S. Barbarossa and G.B. Giannakis. Filterbank Transceivers Optimizing Information Rate in Block Transmission over Dispersive Channels. *IEEE Transactions on Information Theory*, pages 1019–1032, April 1999.

[2] S. Adireddy and L. Tong. Optimal Embedding of Known Symbols. In *34th Annual Conference on Information Sciences and Systems*, March 2000.

[3] N. Al-Dhahir and J.M.Cioffi. Efficiently-Computed Reduced-Parameter Input-Aided MMSE Equalizers for ML Detection : A Unified Approach. *IEEE Trans. Information Theory*, pages 903–915, May 1996.

[4] J. Bingham. Multicarrier modulation for data transmission : An idea whose time has come. *IEEE Communications Magazine*, pages 5–14, May 1990.

[5] T. Cover and J. Thomas. *Elements of Information Theory.* Wiley, 1991.

[6] S. Crozier, D.D. Falconer, and S. Mahmoud. Reduced complexity short-block data detection techniques for fading time-dispersive channels. *IEEE Trans. Vehicular Technology*, pages 255–265, August 1992.

[7] R. Horn and C. Johnson. *Topics in Matrix Analysis.* Cambridge University Press, 1994.

[8] G. Kaleh. Channel Equalization for Block Transmission Systems. *IEEE Journal on Selected Areas in Communications*, pages 110–121, January 1995.

[9] S. Kasturia, J. Aslanis, and J.M. Cioffi. Vector Coding for Partial-Response Channels. *IEEE Transactions on Information Theory*, 36(4):741–762, July 1990.

[10] S. Kay. *Fundamentals of Statistical Signal Processing : Estimation Theory*. Prentice Hall, 1993.

[11] J. Proakis. *Digital Communications*. McGraw Hill, 1995. Third Edition.

[12] A. Ruiz, J.M. Cioffi, and S. Kasturia. Discrete Multiple Tone Modulation with Coset Coding for the Spectrally Shaped Channel. *IEEE Transactions on Communications*, 40(6):1012–1029, June 1992.

[13] J. Seoane, S.K. Wilson, and S. Gelfand. Analysis of intertone and interblock interference in OFDM when the length of the cyclic prefix is shorter than the length of the impulse response of the channel. In *Global Telecommunications Conference*, volume 1, pages 32–36, 1997.

	$0.1 \leq \frac{\nu}{N} < 1$	$0 < \frac{\nu}{N} \leq 0.1$
Low SNR	joint $\mathbf{R}_{xx}$ and $\mathbf{G}$ optimization as in Sec. III–A	set $\mathbf{G} = \left[\begin{array}{cc} \mathbf{I}_\nu & \mathbf{0}_{\nu\times(N-\nu)} \end{array} \right]$ as in DMT optimize $\mathbf{R}_{xx}$ using water–pour procedure [5]
High SNR	set $\mathbf{R}_{xx} = E_x\mathbf{I}_N$ optimize $\mathbf{G}$ as in Sec. III–C	set $\mathbf{G} = \left[\begin{array}{cc} \mathbf{I}_\nu & \mathbf{0}_{\nu\times(N-\nu)} \end{array} \right]$ as in DMT set $\mathbf{R}_{xx} = E_x\mathbf{I}_N$

Table I. Effect of System Parameters on Choice of $\mathbf{R}_{xx}$ and $\mathbf{G}$

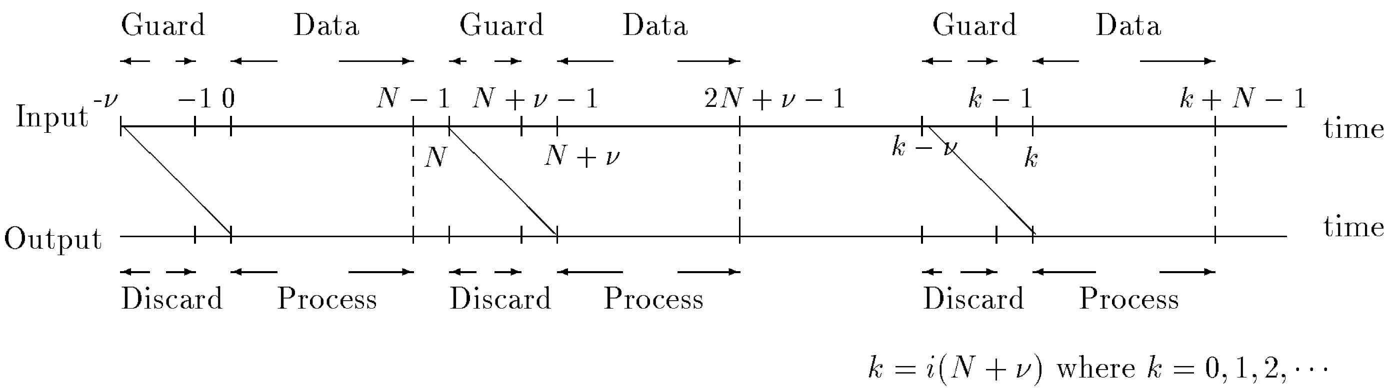

Fig. 1. Role of Guard Sequence in Block Transmission Systems

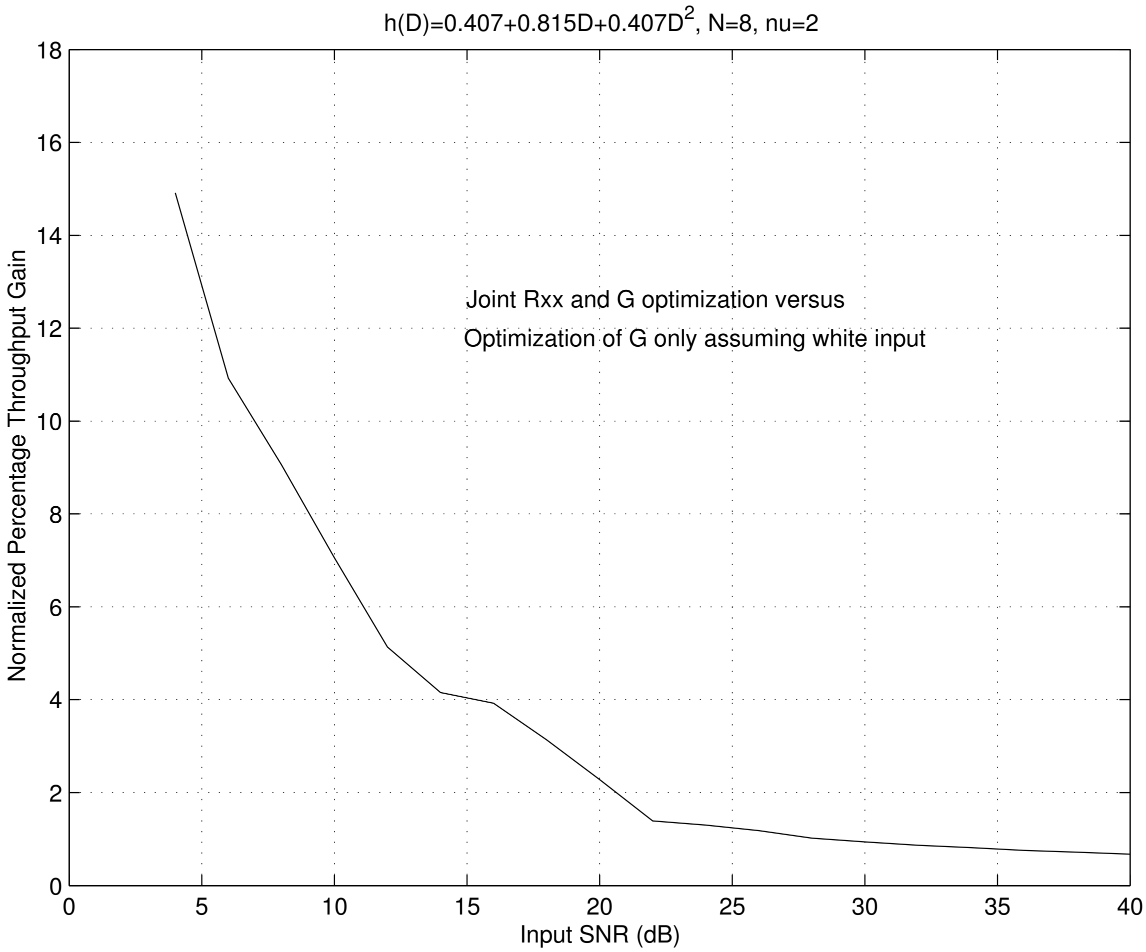

Fig. 2. Normalized Percentage Throughput Gain of the Optimum Scheme of Section III-A over the White-Input Scheme ($\mathbf{R}_{xx} = E_x\mathbf{I}_N$) at Different Input SNR Levels

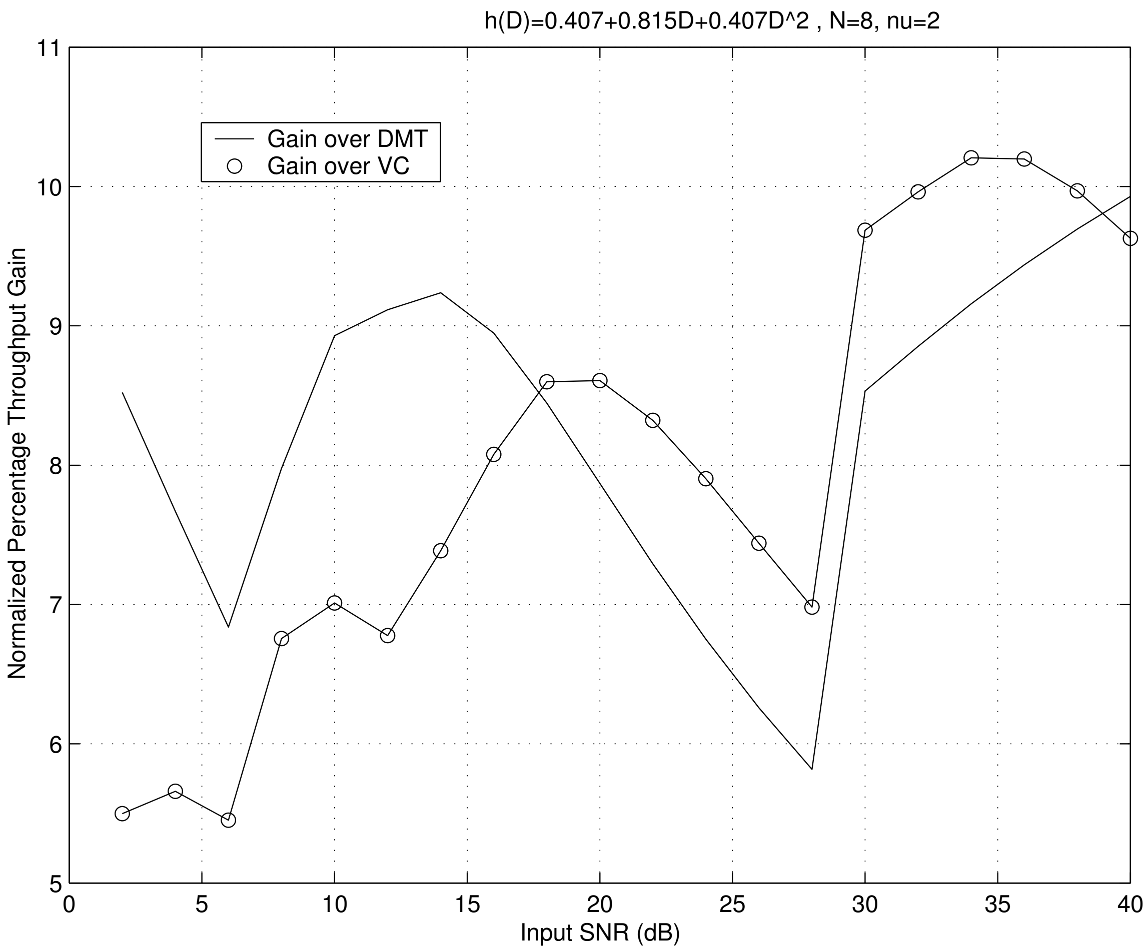

Fig. 3. Normalized Percentage Throughput Gain of Optimum Scheme of Section III-A over VC and DMT

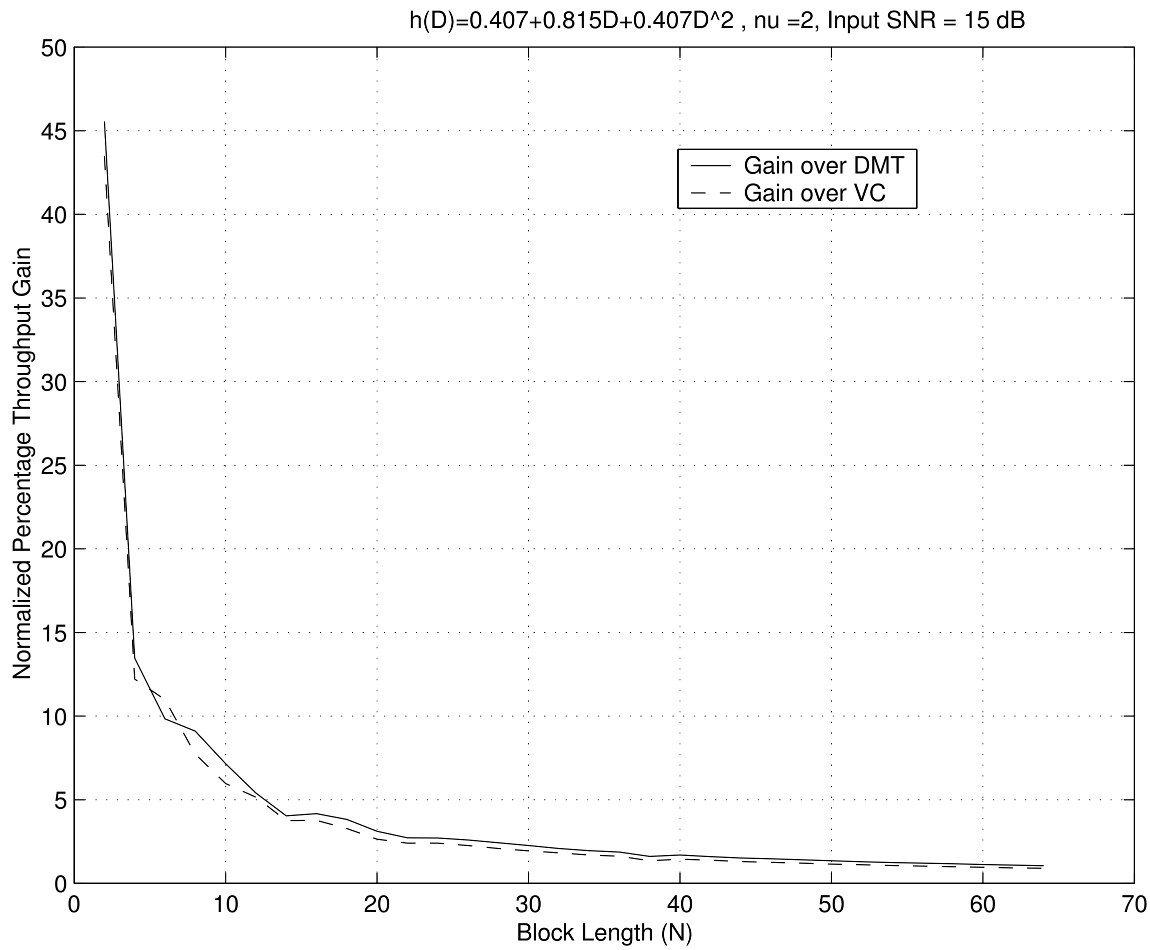

Fig. 4. Normalized Percentage Throughput Gain of Optimum Scheme of Section III-A over VC and DMT as a Function of Block Length

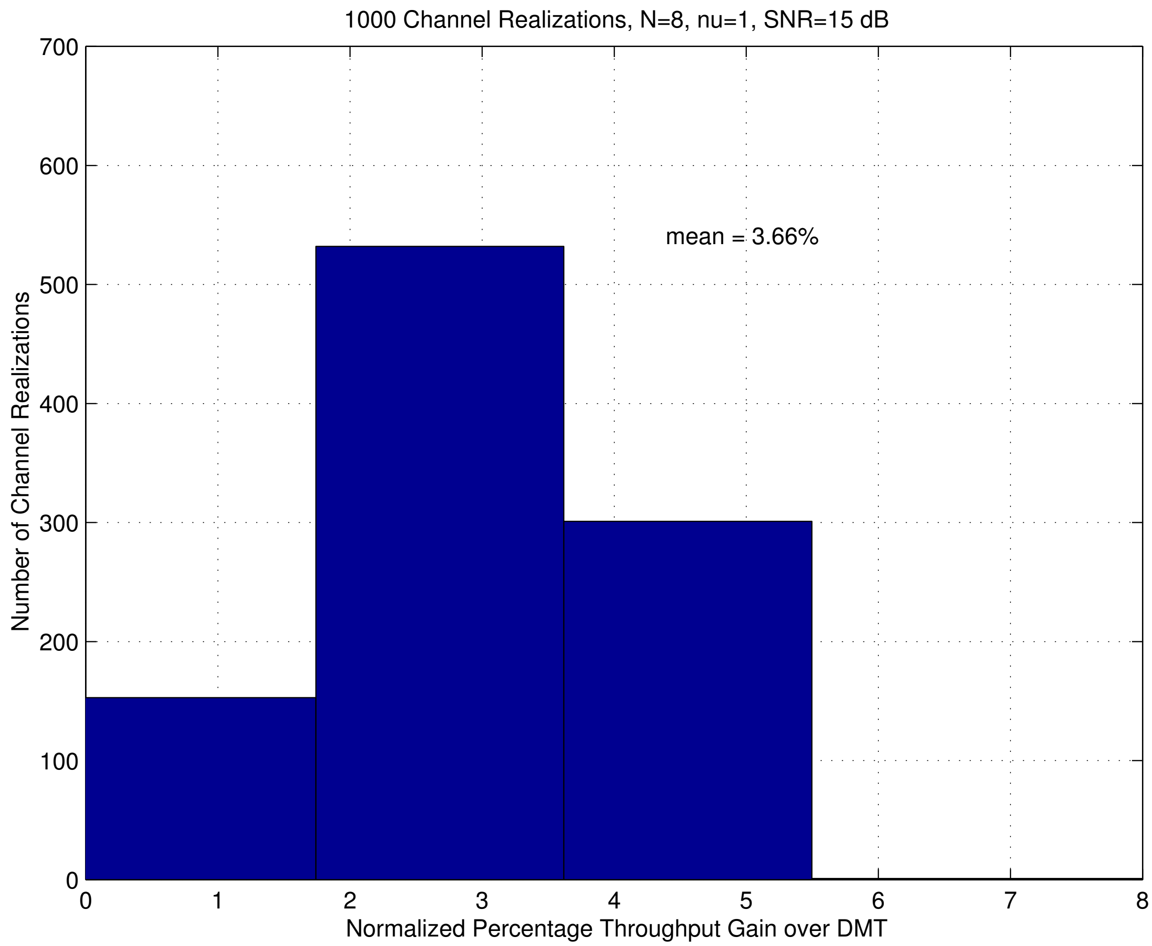

Fig. 5. Normalized Percentage Throughput Gain of Optimum Scheme of Section III-A over DMT for $\nu = 1$, $N = 8$, and $SNR = 15$ dB

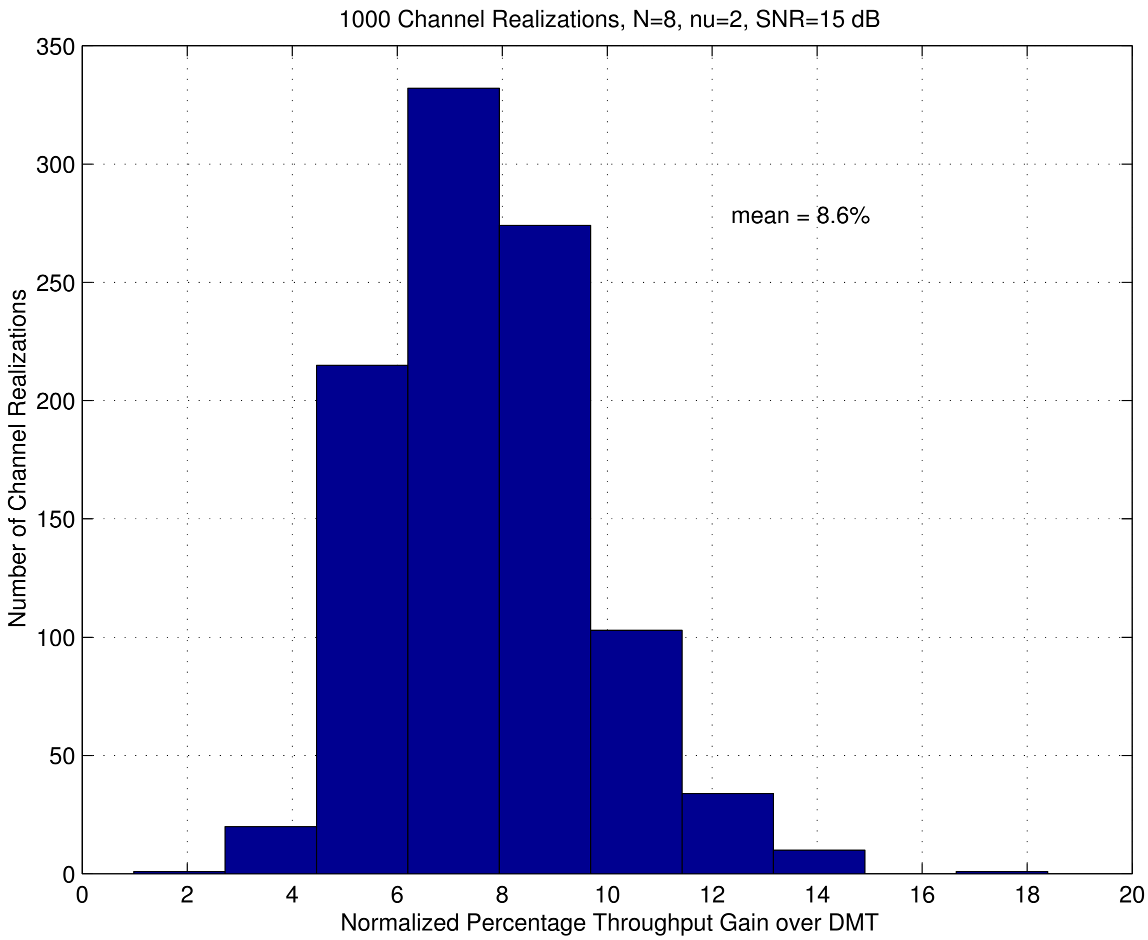

Fig. 6. Normalized Percentage Throughput Gain of Optimum Scheme of Section III-A over DMT for $\nu = 2$, $N = 8$, and $SNR = 15$ dB

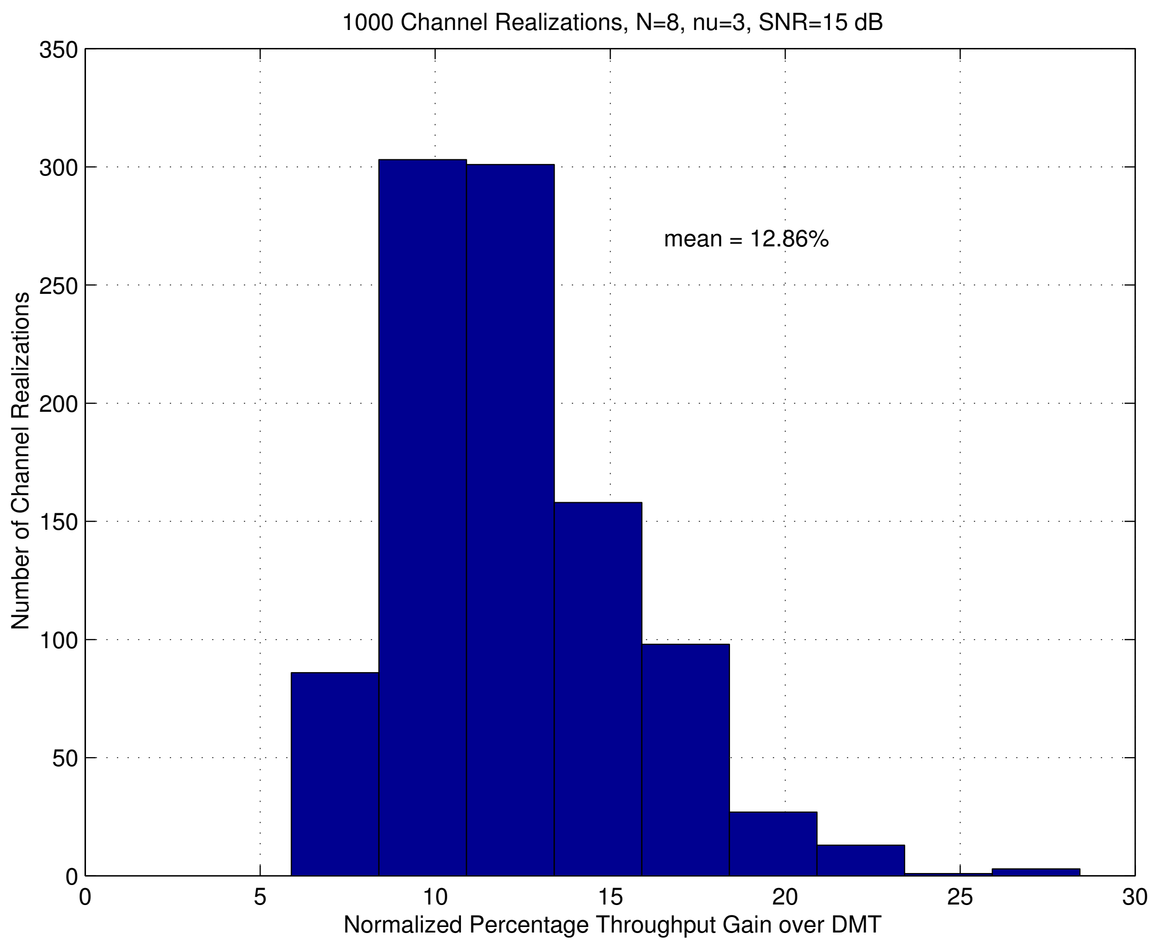

Fig. 7. Normalized Percentage Throughput Gain of Optimum Scheme of Section III-A over DMT for $\nu = 3$, $N = 8$, and $SNR = 15$ dB

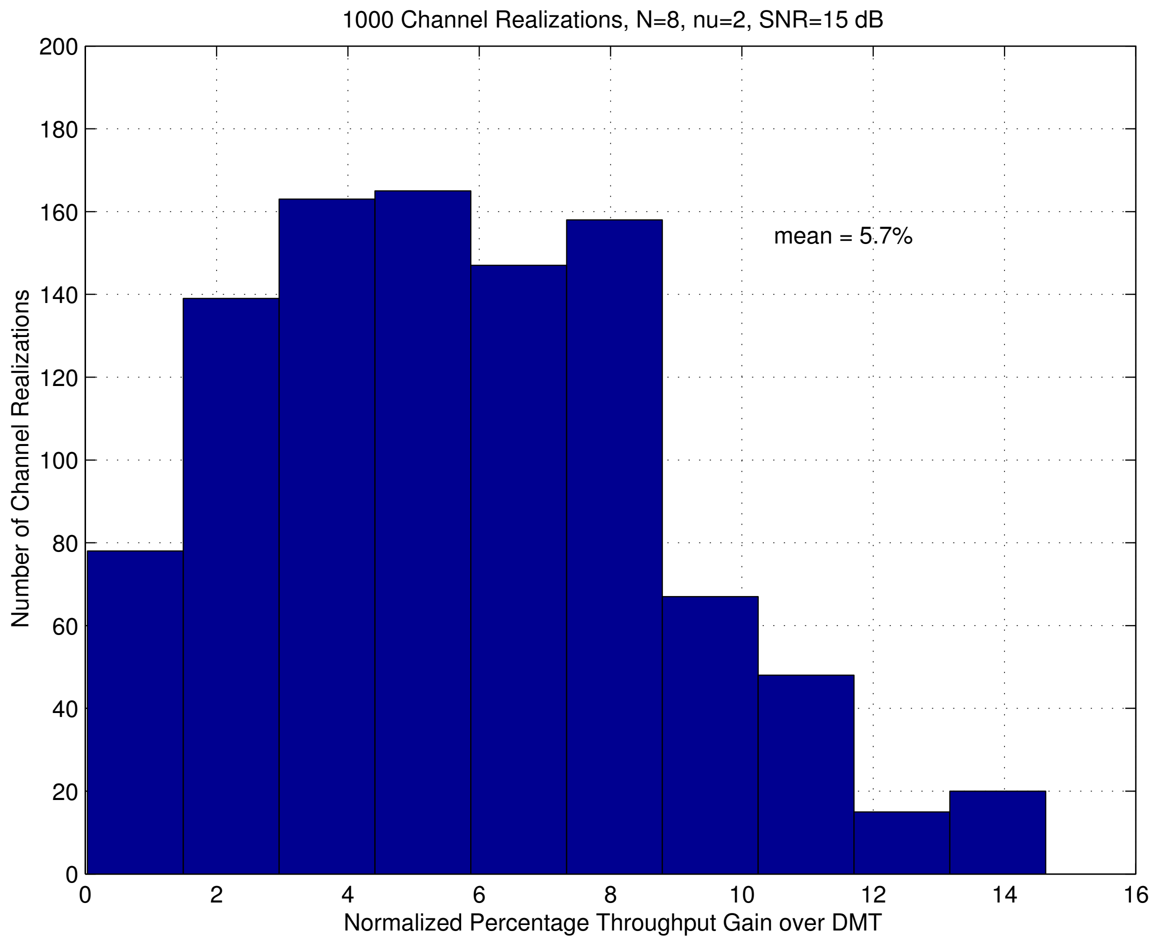

Fig. 8. Normalized Percentage Throughput Gain of Suboptimum Scheme of Section III-B over DMT for $\nu = 2$, $N = 8$, and $SNR = 15$ dB. Both Schemes Use the Optimum $\mathbf{R}_{xx}$ of DMT