23rd International Symposium on Mathematical Theory of Networks and Systems
Hong Kong University of Science and Technology, Hong Kong, July 16-20, 2018

Decision Feedback Based Transceiver Optimization for MIMO Inter-Symbol Interference Channels*

Yi Jiang[1] and Daniel P. Palomar[2] and Mahesh K. Varanasi[3]

***Abstract*— This paper deals with optimizing a transceiver, which consists of a linear time invariant (LTI) matrix filter at the transmitter and a generalized (nonlinear) decision feedback (DF) receiver, for *single-carrier* data transmission over multi-input multi-output (MIMO) inter-symbol interference (ISI) channels. Considering a spatial multiplexing approach for which multiple scalar substreams are transmitted simultaneously, we adopt some cost function $f_0(\{\mathrm{MSE}_i\})$ to measure the overall system performance, where MSE_i's are the mean-squared errors (MSE) of the unquantized estimates at the receiver. Based on majorization theory and the generalized triangular decomposition (GTD), we derive the optimum DF based transceiver which minimizes $f_0(\{\mathrm{MSE}_i\})$ subject to the total input power constraint. It is proven that for *any* cost function f_0 the optimum transmitter is of the same special structure and hence the original complicated matrix optimization problem can be significantly simplified to a problem with scalar variables. Furthermore, if the cost function is specialized to the cases where the composite function $f_0 \circ \exp$ is Schur-convex, then the optimum nonlinear transceiver turns out to be a generalization of the uniform channel decomposition (UCD) scheme; when $f_0 \circ \exp$ is Schur-concave, the optimum nonlinear design degenerates to diagonal transmission, which converts a MIMO ISI channel into multiple decoupled SISO ISI channels.**

***Index Terms*— MIMO transceiver optimization, majorization theory, Schur-convex, matrix spectral factorization.**

I. INTRODUCTION

In digital communications, the channel state information (CSI) can often be made available to the transmitter (CSIT) either through feedback or using the channel reciprocity when time division duplex (TDD) is used. Using full CSI, one may jointly optimize the transmitter and receiver pair for high rate and reliable communication. Much research effort has been placed on mulit-input multi-output (MIMO) transceiver designs since the late 1990s [1][2][3]. Using the majorization theory and the convex optimization theory [4], the authors of [2] established a unifying framework which encompasses the existing linear transceiver designs. A different paradigm of nonlinear transceiver designs was developed based on matrix decomposition algorithms [5][6] and from a channel decomposition perspective [7][8][9]. The nonlinear designs employ a linear transmitter with a (nonlinear) decision feedback (DF) receiver [10]. The designs of the linear and nonlinear transceivers are unified using majorization theory in [11] . Due to the popularity of the orthogonal frequency division modulation (OFDM) technology, which divides an ISI (frequency elective) channel into multiple ISI-free (frequency flat) channel, the endeavors of MIMO transceiver design focus mostly on the ISI-free channels.

In this paper, we consider the transceiver optimization problem for single-carrier data transmission over MIMO ISI channels, where multiple independent SISO substreams are spatially multiplexed and transmitted but without using OFDM. The quality of the substreams is measured by the MSEs of their respective estimates at the receiver side. We establish a theoretical framework of jointly optimizing the transceiver, which consists of a linear time invariant (LTI) matrix filter at the transmitter and a generalized (nonlinear) decision feedback equalizer (DFE) [10] at the receiver, according to some cost function $f_0(\{\mathrm{MSE}_i\})$ subject to the total input power constraint. We first show that for any cost function f_0 which is increasing in each argument, the optimum DF receiver must be the minimum mean squared error DFE (MMSE-DFE). Using the generalized triangular decomposition (GTD) [6] and majorization theory [4], we further prove that for *any* cost function f_0 the optimum transmitter is of the same special structure. Based on this observation, the original complicated matrix optimization problem can be greatly simplified to an optimization problem with scalar-valued variables. Moreover, if the cost function is such that the composite function $f_0 \circ \exp$ is Schur-convex, then the optimum nonlinear transceiver design turns out to be a generalization of the uniform channel decomposition (UCD) scheme proposed in [7]. On the other hand, when $f_0 \circ \exp$ is Schur-concave, the optimum nonlinear design degenerates to diagonal transmission, which converts a MIMO ISI channel into multiple decoupled SISO ISI channels.

II. TRANSCEIVER ARCHITECTURE AND PROBLEM FORMULATION

Consider a MIMO ISI channel with M_t transmit and M_r receive antennas. Assuming pulse shaping filter and Nyquist sampling, the received sampled baseband data are

$$\mathbf{y}_t = \sum_{k=0}^{\infty} \mathbf{H}_k \mathbf{s}_{t-k} + \mathbf{n}_t, \quad t \in \mathbb{Z} \tag{1}$$

where $\{\mathbf{y}_t\}$ is the received M_r-dimensional vector sequence, $\{\mathbf{s}_t\}$ is the transmitted M_t-dimensional vector sequence

*This work is supported in part by NSFC No. 61771005, Hong Kong RGC 16206315 research grant, NSF Grant 1423657 and Qualcomm Gift 24868.

[1]Yi Jiang is with the Dept. of Communication Science and Engineering, Fudan University, Shanghai, China 200438 (yijiang@fudan.edu.cn)

[2]Daniel P. Palomar is with Dept. of ECE, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong (Palomar@ust.hk)

[3] Mahesh K. Varanasi is with the Dept. of ECE, University of Colorado, Boulder, CO 80309-0425, USA (varanasi@colorado.edu)

which is assumed to be stationary and have the average power

$$P_T = \mathbb{E}[\|\mathbf{s}_t\|^2]. \tag{2}$$

Without loss of generality, $\{\mathbf{n}_t\}$ is assumed to be a stationary vector Gaussian process white in both spatial and temporal domain, i.e., $\mathbf{n}_t \sim N(\mathbf{0}, \mathbf{I}_{M_r})$ and $\mathbb{E}[\mathbf{n}_t\mathbf{n}_{t-n}^*] = \delta(n)\mathbf{I}$, where $(\cdot)^*$ is the conjugate transpose, and $\delta(n)$ is the Kronecker delta. The channel response denoted by the matrix sequence $\{\mathbf{H}_k\}$ is known to both the transmitter and receiver and remains fixed within a reasonably long time period.

Using the D-transform

$$\mathcal{H}(D) = \sum_{k=0}^{\infty} \mathbf{H}_k D^k \tag{3}$$

and regarding D as the unit delay operator, i.e., $D^k\mathbf{s}_t = \mathbf{s}_{t-k}$, we can represent (1) by

$$\mathbf{y}_t = \mathcal{H}(D)\mathbf{s}_t + \mathbf{n}_t, \quad t \in \mathbb{Z}. \tag{4}$$

A. Transceiver Architecture

This paper deals with the transceiver architecture consisting of a linear transmitter and a DFE as illustrated in Figure 1. The transmitted sequence $\{\mathbf{s}_t\}$ is obtained by transforming the L-dimensional information sequence $\{\mathbf{x}_t\}$ using a causal LTI matrix filter:

$$\mathbf{s}_t = \sum_{k=0}^{\infty} \mathbf{P}_k\mathbf{x}_{t-k} = \mathcal{P}(D)\mathbf{x}_t, \tag{5}$$

where $\mathbf{P}_k \in \mathbb{C}^{M_t \times L}$ $\forall k$, $\mathcal{P}(D)$ is similarly defined as $\mathcal{H}(D)$ in (3) and the L entries of $\mathbf{x}_t$ are the L independent substreams at time t. We shall see later that the major function of $\mathcal{P}(D)$ is to shape the power spectrum of the input signal in both spatial and frequency domain. Substituting (5) into (4), we have the received signal

$$\mathbf{y}_t = \mathcal{H}(D)\mathcal{P}(D)\mathbf{x}_t + \mathbf{n}_t, \quad t \in \mathbb{Z}. \tag{6}$$

We assume that $\{\mathbf{x}_t\}$ is a stationary vector random process with auto-correlation function $\mathbb{E}[\mathbf{x}_t\mathbf{x}_{t-n}^*] = \mathbf{I}\delta(n)$. Hence

$$\begin{aligned} P_T &= \mathbb{E}\|\mathbf{s}_t\|^2 = \sum_{k=0}^{\infty} \mathrm{Tr}(\mathbf{P}_k\mathbf{P}_k^*) \\ &= \int_0^1 \mathrm{Tr}(\mathcal{P}(e^{j2\pi f})\mathcal{P}^*(e^{j2\pi f}))df, \end{aligned} \tag{7}$$

where the last equality is due to the generalized Paseval's identity. Here we emphasize that

$$\mathcal{P}^*(e^{j2\pi f}) \triangleq [\mathcal{P}(e^{j2\pi f})]^* = \sum_{k=0}^{\infty} \mathbf{P}_k^* e^{-j2\pi fk}. \tag{8}$$

At the receiver side, we apply the DFE to the received data vector sequence $\{\mathbf{y}_t\}$ to recover the information vector sequence $\{\mathbf{x}_t\}$. The DFE consists of a feed-forward filter (FFF), which is an anti-causal LTI MIMO system with D-transform $\mathcal{W}^*(D) = \sum_{i=-\infty}^{0} \mathbf{W}_t^* D^i$, a feed-backward filter (FBF), which is a causal LTI system $\mathcal{B}(D) = \sum_{i=0}^{\infty} \mathbf{B}_i D^i$, and a (nonlinear) quantizer $\mathcal{Q}[\cdot]$ which maps the analog estimates onto the closest constellation points [12]. As the input into the detector, the analog estimate of $\mathbf{x}_t$ is

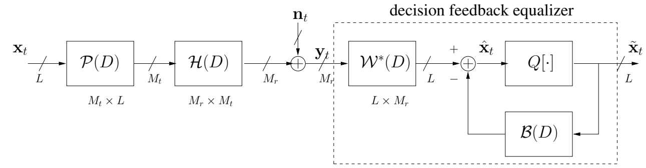


Fig. 1. Scheme of a MIMO Communication system with decision feedback equalizer (DFE) receiver

$$\hat{\mathbf{x}}_t = \mathcal{W}^*(D)\mathbf{y}_t - \mathcal{B}(D)\tilde{\mathbf{x}}_t. \tag{9}$$

Denote $x_{i,t}$ as the ith entry of $\mathbf{x}_t$. The DFE detects signals *successively* in the ordering:

$$\begin{aligned} &\{\ldots, x_{L,t-1} \to x_{L-1,t-1} \to \ldots \to x_{1,t-1} \\ &\to x_{L,t} \to x_{L-1,t} \to \ldots\}, \end{aligned} \tag{10}$$

which is a zig-zag path in the spatial-temporal domain. Corresponding to this detection ordering, $\mathbf{B}_0$, the leading matrix of the series $\mathcal{B}(D)$, should be a strictly upper triangular matrix. Invoking the usual simplifying assumption that the quantizer gives correct estimates of the information sequence, i.e., $\tilde{\mathbf{x}}_t = \mathbf{x}_t$, we obtain that

$$\hat{\mathbf{x}}_t = [\mathcal{W}^*(D)\mathcal{H}(D)\mathcal{P}(D) - \mathcal{B}(D)]\mathbf{x}_t + \mathcal{W}^*(D)\mathbf{n}_t, \tag{11}$$

and the error vector is

$$\mathbf{e}_t \triangleq \hat{\mathbf{x}}_t - \mathbf{x}_t = \mathcal{F}(D)\mathbf{x}_t + \mathcal{W}^*(D)\mathbf{n}_t, \tag{12}$$

where $\mathcal{F}(D) = \mathcal{W}^*(D)\mathcal{H}(D)\mathcal{P}(D) - \mathcal{B}(D) - \mathbf{I}$. The D-transform of the autocorrelation sequence of $\mathbf{e}_t$ is [recall that $\mathbb{E}[\mathbf{n}_t\mathbf{n}_{t-n}^*] = \mathbb{E}[\mathbf{x}_t\mathbf{x}_{t-n}^*] = \mathbf{I}\delta(n)$]

$$\mathcal{E}(D) = \mathcal{F}(D)\mathcal{F}^*(D^{-1}) + \mathcal{W}^*(D)\mathcal{W}(D^{-1}), \tag{13}$$

where $\mathcal{F}^*(D^{-1}) \triangleq [\mathcal{F}(D^{-*})]^* = \sum_{i=0}^{\infty} F_i^* D^{-i}$. By the Wiener-Khinchin Theorem, the MSE matrix is

$$\mathbf{E} \triangleq \mathbb{E}[\mathbf{e}_t\mathbf{e}_t^*] = \int_0^1 \mathcal{E}(e^{j2\pi f})df, \tag{14}$$

and the MSE of the ith substream is

$$\mathrm{MSE}_i \triangleq [\mathbf{E}]_{ii} \quad \text{for} \quad 1 \le i \le L, \tag{15}$$

where $[\mathbf{A}]_{ij}$ stands for the (i,j)th entry of $\mathbf{A}$. Clearly, the MSEs depend on the transmitter $\mathcal{P}(D)$ and the DFE $\mathcal{W}(D)$ and $\mathcal{B}(D)$.

B. Problem Formulation

We study the problem of jointly optimizing the MIMO transmitter and receiver by some global cost function of the MSEs subject to the input power constraint. This problem was studied for linear MIMO transceiver designs in the scenario of ISI-free channel [2]. Here we consider the DF based nonlinear transceiver design in the more general

MIMO-ISI channel. The nonlinear DF design optimization can be represented by the following generic form:

$$\begin{array}{ll} \underset{\mathcal{P}(D),\mathcal{W}(D),\mathcal{B}(D)}{\text{minimize}} & f_0(\{\text{MSE}_i\}) \\ \text{subject to} & \int_0^1 \text{Tr}(\mathcal{P}(e^{j2\pi f})\mathcal{P}^*(e^{j2\pi f}))df \leq P_0, \end{array} \tag{16}$$

where the cost function $f_0(\{\text{MSE}_i\})$ is increasing in each argument and is chosen based on some criterion of practical significance. For example, the cost function may include but not be limited to (i) the sum of MSEs ($f_0(\{\text{MSE}_i\}) = \sum_{i=1}^L \text{MSE}_i$), (ii) the weighted sum of the MSEs ($f_0(\{\text{MSE}_i\}) = \sum_{i=1}^L \text{MSE}_i$), (iii) the product of the MSEs $f_0(\{\text{MSE}_i\}) = \prod_{i=1}^L \text{MSE}_i$, and (iv) the maximal MSE ($f_0(\{\text{MSE}_i\}) = \max_{1\leq i\leq L}\text{MSE}_i$). The paper [13] considered the cost function (i) and (iii).[1] Therefore, (16) includes the problems of [13] as special cases. Moreover, even for the cost functions (i) and (iii), we can obtain better transceiver designs than those obtained in [13] due to the GTD algorithm[6], as we shall discuss later.

C. Optimal Receiver

To solve the seemingly complicated optimization problem (16), we first optimize the receiver, i.e., the FFF $\mathcal{W}(D)$ and the FBF $\mathcal{B}(D)$ as functions of the transmitter $\mathcal{P}(D)$. This result is summarized in the following theorem.

Theorem 2.1: Let

$$\mathcal{P}^*(D^{-1})\mathcal{H}^*(D^{-1})\mathcal{H}(D)\mathcal{P}(D) + \mathbf{I} = \mathcal{R}^*(D^{-1})\mathcal{R}(D). \tag{17}$$

be the matrix spectral factorization where $\mathcal{R}(D) = \sum_{n=0}^\infty \mathbf{R}_n D^n$ and $\mathbf{R}_0$ is upper triangular. Denote $\mathbf{D}_{R_0}$ the diagonal matrix with the same diagonal as $\mathbf{R}_0$. For any FFF and FBF pair $\{\mathcal{W}(D), \mathcal{B}(D)\}$, the MSEs

$$\text{MSE}_i \geq [\mathbf{D}_{R_0}^{-2}]_{ii}, \quad 1 \leq i \leq L, \tag{18}$$

where equality holds if and only if

$$\begin{aligned} &\mathcal{W}(D) = \mathcal{H}(D^{-1})\mathcal{P}(D^{-1})[\mathcal{R}(D^{-1})]^{-1}\mathbf{D}_{R_0}^{-1}, \\ &\mathcal{B}(D) = \mathbf{D}_{R_0}^{-1}\mathcal{R}(D) - \mathbf{I}, \end{aligned} \tag{19}$$

respectively. Moreover, if the equality in (18) is achieved, the MSE matrix $\mathbf{E} = \mathbf{D}_{R_0}^{-2}$ is diagonal.

Proof: The derivation is rather standard and thus is omitted. ■

It is important to note that using the DFE given in (19) the minimum MSEs are achieved simultaneously without incurring tradeoffs among them. Such a DFE is called the MMSE-DFE [12]. According to this theorem, for any cost function $f_0(\{\text{MSE}_i\})$ which is increasing in each argument, the optimum receiver must be the MMSE-DFE. With this observation, the MIMO transceiver optimization problem is significantly simplified:

$$\begin{array}{ll} \underset{\mathcal{P}(D)}{\text{minimize}} & f_0\left(\{[\mathbf{R}_0]_{ii}^{-2}\}\right) \\ \text{subject to} & \mathcal{P}^*(D^{-1})\mathcal{H}^*(D^{-1})\mathcal{H}(D)\mathcal{P}(D) + \mathbf{I} \\ & \qquad = \mathcal{R}^*(D^{-1})\mathcal{R}(D) \\ & \mathcal{R}(D) = \sum_{i=0}^\infty \mathbf{R}_i D^i \\ & \int_0^1 \text{Tr}(\mathcal{P}(e^{j2\pi f})\mathcal{P}^*(e^{-j2\pi f}))df \leq P_0. \end{array} \tag{20}$$

Using the MMSE-DFE, the output signal-to-inference-noise ratios (SINR) of the substreams is related to the MSEs by [10]

$$\text{SINR}_i = \frac{1}{\text{MSE}_i} - 1. \tag{21}$$

Suppose that the transmitted data substreams are Gaussian codes. Then the L substreams effectively pass through L scalar subchannels with mutual information

$$R_i = \log(1+\text{SINR}_i) = -\log \text{MSE}_i = \log[\mathbf{R}_0]_{ii}^2, 1 \leq i \leq L. \tag{22}$$

The MMSE-DFE is information lossless. That is, the sum of the mutual information is

$$\sum_{i=1}^L R_i = -\log\prod_{i=1}^L \text{MSE}_i = \log\prod_{i=1}^L [\mathbf{R}_0]_{ii}^2 \tag{23}$$

$$= \int_0^1 \log\left|\mathbf{I} + \mathcal{H}(e^{j2\pi f})\mathcal{P}(e^{j2\pi f})\mathcal{P}^*(e^{j2\pi f})\mathcal{H}^*(e^{j2\pi f})\right| df. \tag{24}$$

where the right hand side is the mutual information of the MIMO ISI channel.

III. Transmitter Optimization for ISI Channel

In this section, we derive the optimum transmitter according to (20).

A. General Cost Function

Two major hurdles complicate the problem (20): (i) there is no analytical expression for the diagonal of $\mathbf{R}_0$ in terms of the transmitter filter $\mathcal{P}(D)$, and (ii) it is a variational optimization problem as the variable to optimize is a function rather than a finite dimensional vector. However, based on majorization theory and the generalized triangular decomposition (GTD) [6], we circumvent the two hurdles and show that (20) can be solved in a rather simple way. To this end, we first introduce a lemma with respect to the diagonal of $\mathbf{R}_0$. The full proof will be included in the journal version of the paper, and is omitted here due to the space limit.

Lemma 3.1: For the matrix spectral factorization (17), the diagonal of $\mathbf{R}_0$, the leading matrix of $\mathcal{R}(D)$, must satisfy

$$\begin{aligned} &([\mathbf{R}_0]_{11}^2, \ldots, [\mathbf{R}_0]_{LL}^2) \\ \prec_\times \quad &\left\{\exp\left(\int_0^1 \log\left(1+\sigma_{HP,i}^2(e^{j2\pi f})\right)df\right)\right\}_{i=1}^L \end{aligned} \tag{25}$$

where $\sigma_{HP,i}(e^{j2\pi f})$ is the ith largest singular value of $\mathcal{H}(e^{j2\pi f})\mathcal{P}(e^{j2\pi f})$.

Based on Lemma 3.1, we can prove the following theorem of more general interest.

[1] In [13], the considered cost function is the determinant $|\mathbf{E}|$. However, we shall see soon that $\mathbf{E}$ is diagonal if the MMSE-DFE is used. Hence $|\mathbf{E}| = \prod_{i=1}^L \text{MSE}_i$.

Lemma 3.1 is only one-directional. That is, Lemma 3.1 does *not* imply that the squared diagonal of $\mathbf{R}_0$ can be *any* point in the set

$$\left\{\mathbf{x} \in \mathbb{R}_+^L : \mathbf{x} \prec_\times \left\{\exp\left(\int_0^1 \log\left(1+\sigma_{HP,i}^2(e^{j2\pi f})\right) df\right)\right\}_{i=1}^L\right\}. \tag{26}$$

In this sense, the problem (20) and the following are not necessarily equivalent

$$\begin{array}{ll} \underset{\mathcal{P}(D),\{[\mathbf{R}_0]_{ii}^2\}}{\text{minimize}} & f_0\left(\{[\mathbf{R}_0]_{ii}^{-2}\}\right) \\ \text{subject to} & \{[\mathbf{R}_0]_{ii}^2\} \prec_\times \\ & \left\{\exp\left(\int_0^1 \log(1+\sigma_{HP,i}^2(e^{j2\pi f}))df\right)\right\} \\ & \int_0^1 \mathrm{Tr}(\mathcal{P}(e^{j2\pi f})\mathcal{P}^*(e^{-j2\pi f}))df \le P_0. \end{array} \tag{27}$$

Indeed, the minimized cost function in (27) is a lower bound to that in (20). But we shall show soon that this lower bound is achievable.

Let us first consider the solution to (27). We have the following lemma.

Lemma 3.2: The solution to (27) may be set to be $\mathcal{P}(e^{j2\pi f}) = \mathbf{V}_H(e^{j2\pi f})\mathbf{\Sigma}_P(e^{j2\pi f})\mathbf{\Omega}(e^{j2\pi f})$ without loss of optimality, where $\mathbf{V}_H(e^{j2\pi f})$ is the right singular matrix of $\mathcal{H}(e^{j2\pi f})$, $\mathbf{\Sigma}_P(e^{j2\pi f})$ is the diagonal singular matrix of $\mathcal{P}(e^{j2\pi f})$.

Proof: We give the sketch of the proof.

It is not difficult to prove that

$$\begin{array}{l} \sum_{i=1}^k \log(1+\sigma_{HP,i}^2(e^{j2\pi f})) \\ \le \sum_{i=1}^k \log(1+\sigma_{H,i}^2(e^{j2\pi f})\sigma_{P,i}^2(e^{j2\pi f})), \quad 1 \le k \le K, \end{array} \tag{28}$$

where the equality holds if $\mathcal{P}(e^{j2\pi f}) = \mathbf{V}_H(e^{j2\pi f})\mathbf{\Sigma}_P(e^{j2\pi f})\mathbf{\Omega}$. It follows from (28) that

$$\prod_{i=1}^k \exp\left(\int_0^1 \log(1+\sigma_{HP,i}^2(e^{j2\pi f}))df\right) \le$$
$$\prod_{i=1}^k \exp\left(\int_0^1 \log(1+\sigma_{H,i}^2(e^{j2\pi f})\sigma_{P,i}^2(e^{j2\pi f}))df\right), \tag{29}$$
$$\text{for } 1 \le k \le K.$$

Hence for any feasible $\mathcal{P}(D)$, setting$\mathcal{P}(e^{j2\pi f}) = \mathbf{V}_H(e^{j2\pi f})\mathbf{\Sigma}_P(e^{j2\pi f})\mathbf{\Omega}$. relaxes the majorization constraint in (27) to

$$\begin{array}{l} \{[\mathbf{R}_0]_{ii}^2\}_{i=1}^L \prec_\times \\ \left\{\exp\left(\int_0^1 \log(1+\sigma_{H,i}^2(e^{j2\pi f})\sigma_{P,i}^2(e^{j2\pi f}))df\right)\right\}_{i=1}^L, \end{array} \tag{30}$$

and it does not affect the power constraint

$$\int_0^1 \mathrm{Tr}(\mathcal{P}(e^{j2\pi f})\mathcal{P}^*(e^{-j2\pi f}))df \le P_0.$$

Now it is proven that we may set

$$\mathcal{P}(e^{j2\pi f}) = \mathbf{V}_H(e^{j2\pi f})\mathbf{\Sigma}_P(e^{j2\pi f})\mathbf{\Omega}(e^{j2\pi f}) \tag{31}$$

without loss of optimality.

■

Based on the above observation, we can simplify the solution to (27) to be $\mathcal{P}(e^{j2\pi f}) = \mathbf{V}_H(e^{j2\pi f})\mathbf{\Sigma}_P(e^{j2\pi f})\mathbf{\Omega}(e^{j2\pi f})$ where $\mathbf{\Sigma}_P(e^{j2\pi f})$ and $\mathbf{\Omega}(e^{j2\pi f})$ are solved by

$$\begin{array}{ll} \underset{\mathbf{\Sigma}_P(e^{j2\pi f}),\mathbf{\Omega}(e^{j2\pi f})}{\text{minimize}} & f_0\left(\{[\mathbf{R}_0]_{ii}^{-2}\}\right) \\ \text{subject to} & \{[\mathbf{R}_0]_{ii}^2\} \prec_\times \\ & \left\{\exp(\int_0^1 \log(1+\sigma_{H,i}^2(e^{j2\pi f})\sigma_{P,i}^2(e^{j2\pi f}))df)\right\}_{i=1}^L \\ & \sum_{i=1}^K \int_0^1 \sigma_{P,i}^2(e^{j2\pi f})df \le P_0. \end{array} \tag{32}$$

We want to emphasize that $\mathbf{\Omega}(e^{j2\pi f})$ is relevant in (32) because it influences the diagonal of $\mathbf{R}_0$. The following theorem provides further insights on the solution to (32).

Lemma 3.3: As an optimal solution to (32), $\mathbf{\Omega}(e^{j2\pi f})$ is frequency independent, i.e., $\mathbf{\Omega}(e^{j2\pi f}) = \mathbf{\Omega}$, $\forall f \in [0,1)$. Moreover, for any $\{[\mathbf{R}_0]_{ii}^2\}$ that satisfies

$$\{[\mathbf{R}_0]_{ii}^2\} \prec_\times \left\{\exp\left(\int_0^1 \log(1+\sigma_{H,i}^2(e^{j2\pi f})\sigma_{P,i}^2)df\right)\right\}_{i=1}^L, \tag{33}$$

there exists an orthogonal matrix $\mathbf{\Omega}$ such that the matrix spectral factorization yields $\mathbf{R}_0$ with such diagonal elements.

Proof: The proof is straightforward based on the GTD theorem [6] and is omitted here. ■

We remark that Lemma 3.3 actually shows that the converse direction of Lemma 3.1 is also true, i.e., the squared diagonal of $\mathbf{R}_0$ can be *any* point in the set (26). Therefore, the optimization problem (20) is equivalent to (27) and (32).

As the final step of the derivation, we give an observation with regard to the frequency power allocation. Denote p_i as the total power allocated to the ith eigen-subchannel, $1 \le i \le K$. The optimum power distribution $\sigma_{P,i}^2(e^{j2\pi f})$ must be the solution to the following

$$\begin{array}{ll} \underset{\sigma_{P,i}^2(e^{j2\pi f})}{\text{maximize}} & \int_0^1 \log(1+\sigma_{H,i}^2(e^{j2\pi f})\sigma_{P,i}^2(e^{j2\pi f}))df \\ \text{subject to} & \int_0^1 \sigma_{P,i}^2(e^{j2\pi f})df = p_i. \end{array} \tag{34}$$

This is because maximizing $\int_0^1 \log(1 + \sigma_{H,i}^2(e^{j2\pi f})\sigma_{P,i}(e^{j2\pi f}))df$ relaxes the multiplicative-majorization constraint of (32) to the maximal extent. The solution to (34) is the standard waterfilling power allocation [14]

$$\sigma_{P,i}^2(e^{j2\pi f}) = \left(\mu_i - \frac{1}{\sigma_{H,i}^2(e^{j2\pi f})}\right)^+,$$

where $(x)^+ = \max\{x,0\}$ and μ_i is chosen such that $\int_0^1 \left(\mu_i - \frac{1}{\sigma_{H,i}^2(e^{j2\pi f})}\right)^+ df = p_i$.

At this point, we have transformed the optimization problem (20) to the most simplified form, which is summarized in the following theorem.

Theorem 3.4: Let

$$\mathcal{H}(e^{j2\pi f}) = \mathbf{U}_H(e^{j2\pi f})\mathbf{\Sigma}_H(e^{j2\pi f})\mathbf{V}_H^*(e^{j2\pi f}) \tag{35}$$

be the SVD where the diagonal elements of $\mathbf{\Sigma}_H(e^{j2\pi f})$ are $\sigma_{H,i}(e^{j2\pi f})$. Suppose the following problem

$$\begin{array}{ll} \underset{\boldsymbol{\mu},\ \{[\mathbf{R}_0]_{ii}\}}{\text{minimize}} & f_0\left(\{[\mathbf{R}_0]_{ii}^{-2}\}\right) \\ \text{subject to} & \{[\mathbf{R}_0]_{ii}^2\} \prec_\times \\ & \left\{\exp(\int_0^1 \left(\log(\mu_i\sigma_{H,i}^2(e^{j2\pi f}))\right)^+ df)\right\}_{i=1}^L \\ & \sum_{i=1}^K \int_0^1 \left(\mu_i - \frac{1}{\sigma_{H,i}^2(e^{j2\pi f})}\right)^+ df \le P_0, \end{array} \tag{36}$$

has the optimal solution μ_i and $[\mathbf{R}_0]_{ii}$ for $1 \le i \le L$. Then the optimal precoder has the SVD $\mathcal{P}(e^{j2\pi f}) = \mathbf{V}_H(e^{j2\pi f})\mathbf{\Sigma}_P(e^{j2\pi f})\mathbf{\Omega}$, where $\mathbf{\Sigma}_P(e^{j2\pi f})$ has diagonal $\sigma_{P,i}^2(e^{j2\pi f}) = \left(\mu_i - \frac{1}{\sigma_{H,i}^2(e^{j2\pi f})}\right)^+$ and $\mathbf{\Omega}$ is obtained by applying the GTD to the matrix $\mathbf{\Phi} = \mathbf{Q}\mathbf{R}_0\mathbf{\Omega}^T$ so that the diagonal entries of $\mathbf{R}_0$ are the optimal solution to (36). Here the diagonal matrix $\mathbf{\Phi}$ is defined as

$$[\mathbf{\Phi}]_{ii} \triangleq \exp\left(\frac{1}{2}\int_0^1 \log(1+\sigma_{H,i}^2(e^{j2\pi f})\sigma_{P,i}^2(e^{j2\pi f}))df\right). \tag{37}$$

B. Schur-convex and Schur-concave Cost Functions

Now we further specialize the cost function to the case where f_0 is increasing in each argument and the composite function $f_0 \circ \exp : \mathbb{R}^L \to \mathbb{R}$ is either Schur-convex or Schur-concave, where the composite function is defined as

$$f_0 \circ \exp(\mathbf{x}) \triangleq f_0(e^{x_1}, e^{x_2}, \ldots, e^{x_L}). \tag{38}$$

Many cost functions of interest can be categorized into such functions. The following are some examples.

1) Minimization of the sum of MSEs: The cost function is

$$f_0(\{\text{MSE}_i\}) = \sum_{i=1}^L \text{MSE}_i, \tag{39}$$

which is both Schur-concave and Schur-convex. The composite function $f_0 \circ \exp(\mathbf{x}) = \sum_i e^{x_i}$ is Schur-convex, since e^x is a convex function.

2) Minimization of the exponentially weighted product of MSEs: The cost function is

$$f_0(\{\text{MSE}_i\}) = \prod_{i=1}^L \text{MSE}_i^{\alpha_i}. \tag{40}$$

Without loss of generality, it is assumed that $0 < \alpha_1 \le \ldots \le \alpha_L$. The composite function $f_0 \circ \exp$ is

$$f_0 \circ \exp(\mathbf{x}) = \exp\left(\sum_{i=1}^L \alpha_i x_i\right). \tag{41}$$

It is easy to prove that $\sum_{i=1}^L \alpha_i x_i$ (assuming $\alpha_i \le \alpha_{i+1}$) is a Schur-concave function on $\mathcal{D}_L \triangleq \{\mathbf{x} \in \mathbb{R}^L : x_1 \ge \ldots x_L\}$, so is $\exp\left(\sum_{i=1}^L \alpha_i x_i\right)$.

3) Maximization of the product of MSEs: The objective function to maximize is $\prod_{i=1}^L \text{MSE}_i$. The composite function $f_0 \circ \exp$ is

$$f_0 \circ \exp(\mathbf{x}) = \exp\left(\sum_{i=1}^L x_i\right). \tag{42}$$

Since $\sum_{i=1}^L x_i$ is both Schur-convex and Schur-concave, so is $\exp\left(\sum_{i=1}^L x_i\right)$.

Such specialization leads to an exceedingly simple solution to (20) as shown in the following theorem.

Theorem 3.5: An optimal solution $\mathcal{P}(e^{j2\pi f})$ of the problem (20), where $f_0 : \mathbb{R}^L \to \mathbb{R}$ is a function increasing in each argument, can be characterized as follows:

- If $f_0 \circ \exp$ is Schur-concave on $\mathcal{D}_L \triangleq \{\mathbf{x} \in \mathbb{R}^L : x_1 \ge \ldots \ge x_L\}$, then
$$\mathcal{P}(e^{j2\pi f}) = \mathbf{V}_H(e^{j2\pi f})\mathbf{\Sigma}_P(e^{j2\pi f}), \tag{43}$$
where the diagonal elements of $\mathbf{\Sigma}_P(e^{j2\pi f})$ are $\sigma_{P,i}(e^{j2\pi f}) = \left(\mu_i - \frac{1}{\sigma_{H,i}^2(e^{j2\pi f})}\right)^+$ for $1 \le i \le K$ with μ_i being obtained by solving
$$\begin{array}{ll} \underset{\boldsymbol{\mu}}{\text{minimize}} & f_0\left(\left\{\exp\left(-\int_0^1 \left(\log(\mu_i\sigma_{H,i}^2(e^{j2\pi f}))\right)^+ df\right)\right\}\right) \\ \text{subject to} & \sum_{i=1}^K \int_0^1 \left(\mu_i - \frac{1}{\sigma_{H,i}^2(e^{j2\pi f})}\right)^+ df \le P_0. \end{array} \tag{44}$$
- If $f_0 \circ \exp$ is Schur-convex on $\mathbb{R}^L$, then
$$\mathcal{P}(e^{j2\pi f}) = \mathbf{V}_H(e^{j2\pi f})\mathbf{\Sigma}_P(e^{j2\pi f})\mathbf{\Omega} \tag{45}$$
where $\mathbf{\Sigma}_P(e^{j2\pi f})$ is obtained via standard waterfilling power allocation
$$\sigma_{P,i}^2(e^{j2\pi f}) = \left(\mu - \frac{1}{\sigma_{H,i}^2(e^{j2\pi f})}\right)^+, \quad 1 \le i \le K, \tag{46}$$
with μ being chosen such that
$$\sum_{i=1}^K \int_0^1 \left(\mu - \frac{1}{\sigma_{H,i}^2(e^{j2\pi f})}\right)^+ df = P_0, \tag{47}$$
and the unitary matrix $\mathbf{\Omega}$ is obtained from the geometric mean decomposition (GMD)[5]
$$\mathbf{\Phi} = \mathbf{Q}\mathbf{R}_0\mathbf{\Omega}^T \tag{48}$$
where $\mathbf{\Phi}$ is diagonal with $\mathbf{\Phi}]_{ii} = \exp\left(\frac{1}{2}\int_0^1 \left(\log(\mu\sigma_{H,i}^2(e^{j2\pi f}))\right)^+ df\right)$, $1 \le i \le L$.

Proof: The proof will be given in the journal version of this paper. ∎

For the case that $f_0 \circ \exp$ is Schur-concave, the transmitter matrix filter $\mathcal{P}(e^{2\pi f}) = \mathbf{V}_H(e^{j2\pi f})\mathbf{\Sigma}_P(e^{j2\pi f})$ transforms the MIMO channel into

$$\mathcal{H}(e^{j2\pi f})\mathcal{P}(e^{j2\pi f}) = \mathbf{U}_H(e^{j2\pi f})\mathbf{\Sigma}_H(e^{j2\pi f})\mathbf{\Sigma}_P(e^{j2\pi f}). \tag{49}$$

Note that this effective MIMO channel has orthogonal columns at any frequency. Hence

$$\begin{array}{rl} & \mathcal{P}^*(D^{-1})\mathcal{H}^*(D^{-1})\mathcal{H}(D)\mathcal{P}(D) \\ = & \mathbf{\Sigma}_P^*(D^{-1})\mathbf{\Sigma}_H^*(D^{-1})\mathbf{\Sigma}_H(D)\mathbf{\Sigma}_P(D) \end{array} \tag{50}$$

is diagonal and the matrix spectral factorization

$$\mathcal{P}^*(D^{-1})\mathcal{H}^*(D^{-1})\mathcal{H}(D)\mathcal{P}(D) + \mathbf{I} = \mathcal{R}^*(D^{-1})\mathcal{R}(D)$$

yields $\mathcal{R}(D) = \sum_{i=0}^{\infty} \mathbf{R}_i D^i$ with diagonal $\mathbf{R}_i$ for $\forall i$. The MMSE-DFE given in (19) is

$$\mathcal{W}(D) = \mathbf{U}_H(D^{-1})\mathbf{\Sigma}_H(D^{-1})\mathbf{\Sigma}_P(D^{-1})[\mathcal{R}(D^{-1})]^{-1}\mathbf{D}_{R_0}^{-1}. \tag{51}$$

It is ready to verify that

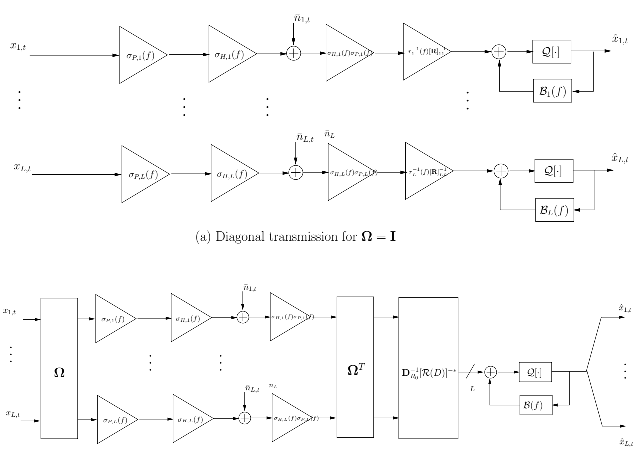
$$\begin{aligned} & \mathcal{W}^*(D)\mathcal{H}(D)\mathcal{P}(D) \\ = \ & \mathbf{D}_{R_0}^{-1}[\mathcal{R}^*(D^{-1})]^{-1}\mathbf{\Sigma}_P^*(D^{-1})\mathbf{\Sigma}_H^*(D^{-1})\mathbf{\Sigma}_H(D)\mathbf{\Sigma}_P(D) \\ = \ & \mathbf{D}_{R_0}^{-1}(\mathcal{R}(D) - [\mathcal{R}^*(D^{-1})]^{-1}). \end{aligned} \tag{52}$$

From the above equation, we see that the transmitter matrix filter and the FFF convert the MIMO ISI channel into K diagonal SISO ISI channels, and the FBF, which consists of L decoupled SISO decision feedback filters, removes the ISI. Figure 2(a) illustrates the architecture of the transceiver, where the function block $r_i^{-1}(f)[\mathbf{R}]_{ii}^{-1}$ corresponds to the item $[\mathcal{R}(D^{-1})]^{-1}\mathbf{D}_{R_0}^{-1}$ in (51).

For *any* cost function f_0 such that $f_0 \circ \exp$ is Schur-convex, the optimum transceiver design is the same! That is, the transmitting matrix filter is a concatenation of $\mathbf{U}_H(D)\mathbf{\Sigma}_P(D)$, which applies capacity-achieving waterfilling power allocation in both spatial and frequency domain, and a frequency-independent unitary matrix which rotates the channel such that the DF receiver, which has the FFF

$$\mathcal{W}(D) = \mathbf{U}_H(D^{-1})\mathbf{\Sigma}_H(D^{-1})\mathbf{\Sigma}_P(D^{-1})\mathbf{\Omega}[\mathcal{R}(D^{-1})]^{-1}\mathbf{D}_{R_0}^{-1}, \tag{53}$$

yields L identical subchannels. (Note that $[\mathcal{R}(D)]$ given in (53), which is non-diagonal, is different from that in (51), which is diagonal.) This is actually an extension of the UCD scheme [7] to the MIMO ISI channel. Figure 2(b) illustrates the architecture of this transceiver.



(a) Diagonal transmission for $\mathbf{\Omega} = \mathbf{I}$

(b) Nondiagonal transmission (diagonal + rotation)

Fig. 2. Scheme of a MIMO Communication system with decision feedback equalizer (DFE) receiver

IV. Conclusion

In this paper, we have studied optimizing the decision feedback (DF) based nonlinear transceivers for data transmission over multi-input multi-output (MIMO) inter-symbol interference (ISI) channels. Adopting some cost function f_0 of the MSEs of the multiplexed substreams, we establish a theoretic framework for transceiver optimization, which consists of a linear time invariant (LTI) matrix filter at the transmitter and a generalized (nonlinear) decision feedback equalizer (DFE) at the receiver, subject to the total input power constraint. Under the mild assumption that f_0 is increasing in MSEs, the optimum DFE receiver is the MMSE-DFE. We further prove that using the MMSE-DFE as the receiver, the optimum transmitter is of the same special structure for *any* cost function f_0, which applies power allocation in both spatial and frequency domain before a spatial rotation. Hence the original complicated matrix optimization problem can be significantly simplified. Moreover, for the cost function such that the composite $f_0 \circ \exp$ is Schur-convex, then the optimum nonlinear transceiver design is a generalized form of the uniform channel decomposition (UCD) scheme; when $f_0 \circ \exp$ is Schur-concave, the optimum nonlinear design degenerates to diagonal transmission.

References

[1] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filter-bank precoders and equalizers part i: Unification and optimal designs," *IEEE Transactions on Signal Processing*, vol. 47, pp. 1988–2006, July 1999.

[2] D. Palomar, J. Cioffi, and M. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Transactions on Signal Processing*, vol. 51, pp. 2381–2401, September 2003.

[3] Y. Jiang, J. Li, and W. Hager, "Joint transceiver design for MIMO communications using geometric mean decomposition," *IEEE Transactions on Signal Processing*, vol. 53, pp. 3791 – 3803, October 2005.

[4] A. Marshall and I. Olkin, *Inequalities: Theory of Majorization.* New York: Academic, 1979.

[5] Y. Jiang, W. Hager, and J. Li, "The geometric mean decomposition," *Linear Algebra and Its Applications*, vol. 396, pp. 373–384, February 2005.

[6] Y. Jiang, W. Hager, and J. Li, "The generalized triangular decomposition," *Mathematics of Computation*, accepted, Nov. 2006.

[7] Y. Jiang, J. Li, and W. Hager, "Uniform channel decomposition for MIMO communications," *IEEE Transactions on Signal Processing*, vol. 53, pp. 4283 – 4294, November 2005.

[8] Y. Jiang, W. Hager, and J. Li, "Tunable channel decomposition for MIMO communications using channel state information," *IEEE Transactions on Signal Processing*, vol. 54, pp. 4405 – 4418, November 2006.

[9] F. Xu, T. Davidson, J. K. Zhang, and K. M. Wong, "Design of block transceivers with decision feedback detection," *IEEE Transactions on Signal Processing*, vol. 54, pp. 965–978, March 2006.

[10] J. M. Cioffi and G. D. Forney, "Generalized decision-feedback equalization for packet transmission with ISI and Gaussian noise," *Communication, Computation, Control and Signal Processing,* A. Paulraj, V. Roychowdhury and C. Schaper, Eds. Boston, MA: Kluwer, 1997, ch.4, pp. 79-127.

[11] D. P. Palomar and Y. Jiang, *MIMO Transceiver Design via Majorization Theory.* Hanover, MA, USA: Now Publishers Inc., 2007.

[12] A. Duel-Hallen, "Equalizers for multiple input/multiple output channels and pam systems with cyclostationary input sequences," *IEEE Journal on Selected Areas In Communications*, vol. 10, pp. 630–639, April 1992.

[13] J. Yang and S. Roy, "Joint transmitter-receiver optimization for multi-input multi-output systems with decision feedback," *IEEE Transactions on Information Theory*, vol. 40, pp. 1334–1347, September 1994.

[14] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* John Wiley & Sons, Inc, 1991.