Robust Constrained-Optimization-Based Linear Receiver for High-Rate MIMO-OFDM Against Channel Estimation Errors

Chih-Yuan Lin, Jwo-Yuh Wu, *Member, IEEE*, and Ta-Sung Lee, *Senior Member, IEEE*

***Abstract*—We consider multi-input multi-output–orthogonal frequency division multiplexing (MIMO-OFDM) transmission in a scenario that the adopted cyclic-prefix (CP) length is shorter than the channel delay spread for boosting data rate and, moreover, the channel parameters are not exactly known but are estimated using the least-squares (LS) training technique. By exploiting the receiver spatial resource, we propose a constrained-optimization-based linear equalizer which can mitigate inter-symbol interference and inter-carrier interference incurred by insufficient CP interval, and is robust against the net detrimental effects caused by channel estimation errors. The optimization problem is formulated in an equivalent unconstrained generalized-sidelobe-canceller (GSC) setup. The channel parameter error is explicitly incorporated into the constraint-free GSC system model through the perturbation technique; this allows us to exploit the presumed LS channel error property for deriving a closed-form solution, and can also facilitate an associated analytic performance analysis. A closed-form approximate signal-to-interference-plus-noise ratio (SINR) expression for the proposed robust scheme is given, and an appealing formula of the achievable SINR improvement over the nonrobust counterpart is further specified. Our analytic results bring out several intrinsic features of the proposed solution. Simulation study confirms the effectiveness of the proposed method and corroborates the predicted SINR results.**

***Index Terms*—Constrained optimization, generalized sidelobe canceller, interference suppression, least-squares channel estimation, multi-input multi-output (MIMO), orthogonal frequency division multiplexing (OFDM), perturbation analysis.**

I. Introduction

A. Motivation and Paper Contributions

ORTHOGONAL frequency division multiplexing (OFDM) combined with multiple transmit and receive antennas, aka, multi-input multi-output (MIMO) OFDM, has become a key communication technique over frequency-selective channels [21], [29]. At the transmitter, the OFDM modulator appends at each symbol block head a cyclic prefix (CP), with length no shorter than the channel delay spread, for removing inter-symbol interference (ISI). The insertion of CP, however, will lower the effective data rate. A commonly used approach to rate boost in MIMO-OFDM is thus to conserve the length of CP, and rely on sophisticated receiver design for combating the residual ISI as well as the induced inter-carrier interference (ICI). Typical such proposals include the time-domain channel shortening mechanism [1] and the frequency-domain per-tone equalization scheme [13]. The former resorts to shortening filter for "squeezing" the composite channel memory within the CP interval, thus limiting the ISI effect; the latter, on the other hand, aims for direct ISI and ICI suppression in the frequency domain. Both of the above methods assume perfect channel information is available at the receiver. However, channel parameter mismatch due to imperfect estimation is inevitable in practice, and can further impair the system performance. To the best of our knowledge, robust receiver design for high-rate MIMO-OFDM against channel estimation errors remains scarce in the literature.

It is well known that multiple receive antennas can provide an array gain for interference rejection [32]. This thus motivates us to leverage the spatial resource for combating ISI-ICI effect in MIMO-OFDM when a limited CP is used for trading high-rate performance. This paper proposes one such solution via a linear constrained-optimization approach [7], assuming that the channel parameters are not exactly known but are estimated via the commonly used least-squares (LS) training technique [2], [19]. The proposed method relies on the block system model for MIMO-OFDM and exploits the inherent joint space-frequency degrees-of-freedom. The adopted interference mitigation mechanism is a simple linear weighting matrix; ISI and ICI are suppressed by minimizing the average power of the filtered interference, while a linear constraint is imposed for extracting the desired signals through maximal ratio combining. To tackle the impact of channel estimation errors, one natural strategy in the considered scenario is to model the channel mismatch as a random variable and exploit the presumed LS channel error characteristic to derive a solution. Toward this end, we resort to the generalized sidelobe canceller (GSC) principle [8], [25], [32] to reformulate the constrained-optimization problem into an equivalent *unconstrained* framework. The constraint-free GSC setup allows us to explicitly track each corrupted signal component resulting from imperfect channel estimation, and in turn leads to a very natural cost function for joint interference and channel error mitigation. We further leverage the perturbation technique [15], [37] to incorporate the channel parameter

Manuscript received August 23, 2006; revised August 24, 2006. This work was supported in part by the National Science Council of Taiwan under Grant NSC 96-2752-E-002-009, in part by the MediaTek Research Center at National Chiao Tung University, Taiwan, and in part by the Ministry of Education of Taiwan under the MoE ATU Program. Part of this paper was presented at the 49th IEEE Global Telecommunications Conference, San Francisco, CA, November 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Erchin Serpedin.

The authors are with the Department of Communication Engineering, National Chiao Tung University, Hsinchu 300, Taiwan, R.O.C. (e-mail: cylin.cm90g@nctu.edu.tw; jywu@cc.nctu.edu.tw; tslee@mail.nctu.edu.tw).

Digital Object Identifier 10.1109/TSP.2007.893206

deviation into the solution equation; the optimal weighting matrix is then obtained by invoking the error characteristics of the LS channel estimate. The proposed optimal solution can mitigate the aggregate impacts due to channel estimation errors like signal leakage and other background parameter perturbation effects. Further analysis reveals that signal leakage turns out to be a dominant factor and a suboptimal solution, in the form of diagonal loading (DL), can attain a near-optimal performance. Also relying on the perturbation technique, we then derive a closed-form signal-to-interference-plus-noise ratio (SINR) expression associated with the suboptimal DL scheme; due to the near-optimal nature of the DL solution, the established result can well predict the actual SINR tendency attained by the optimal one. Our analytic formula can further quantify the achievable average SINR improvement over the nonrobust GSC weight (i.e., the one derived under exact channel knowledge). In particular, we provide a closed-form expression of the SINR increment, and based on which several key features regarding the proposed robust solution can be inferred. Simulation study confirms the effectiveness of the proposed robust solution and also corroborates the presented SINR results.

B. Connection to Previous Works

Robust constrained-optimization based linear receiver design against channel estimation errors has also been addressed in the context of multiuser communication [22], [23], [24], [38]. By modeling the channel mismatch as a random variable with known statistical characteristics, the probability-constrained optimization approach [22], [23] exploits the Gaussianality assumption on the estimation error, and the solution is obtained through linear [22] or nonlinear [23] programming. Although the proposed method in this paper adopts a similar stochastic setting, as we will see it relies entirely on the first- and second-order error statistics but not on the underlying assumption on error distribution. In [24] and [38], the channel error is, on the other hand, treated as a "deterministic" perturbation. Based on a min-max type formulation, the optimization problem in [24] is solved by using the convex programming technique; the solution in [38] admits a DL form, with the optimal loading factor determined via an iterative procedure. We note that the deterministic formulation of model error is also used in robust beamformer design [14], [18], [26], [33], in which exact statistical characterization of model uncertainty due to, e.g., array calibration error, unknown antenna coupling effect, etc., is difficult (or even impossible) to track.

C. Paper Organization and Notation List

The rest of this paper is organized as follows. Section II reviews the MIMO-OFDM system model and the LS channel estimation technique. Section III presents the multichannel system representation and shows the GSC-based ISI-ICI suppression scheme under perfect channel knowledge assumption. The proposed robust GSC filter is derived in Section IV; the related SINR performance analysis is provided in Section V. Section VI discusses the algorithm complexity. Section VII introduces an alternative formulation of the proposed approach. Section VIII is the simulation results. Finally, Section IX is the conclusion.

Notation List: Let $\mathbb{C}^m$ and $\mathbb{C}^{m\times n}$ be the sets of m-dimensional complex vectors and $m\times n$ complex matrices. Denote by $(\cdot)^T, (\cdot)^*$, and $(\cdot)^H$, respectively, the transpose, the complex conjugate, and the complex conjugate transpose. $\mathbf{I}_m$ and $\mathbf{0}_m$ denote the $m\times m$ identity and zero matrices; $\mathbf{0}_{m\times n}$ is the $m\times n$ zero matrix. For $\mathbf{x}\in\mathbb{C}^m$, let $\mathrm{diag}\{\mathbf{x}\}$ be the $m\times m$ diagonal matrix with the elements of $\mathbf{x}$ on the main diagonal. The notation $E\{y\}$ stands for the expected value of the random variable y. For $\mathbf{x}\in\mathbb{C}^m$ and $\mathbf{X}\in\mathbb{C}^{m\times n}$, denote by $\|\mathbf{x}\|$ the vector two-norm and $\|\mathbf{X}\|$ the matrix Frobenius norm. The symbols $\odot$ and $\otimes$ stand for the matrix Hadamard product ([10], p-298) and the matrix Kronecker product ([10], p-242), respectively.

II. Preliminary

A. System Model and Basic Assumptions

We consider the discrete-time baseband model of a MIMO-OFDM system with Q subcarriers, N transmit antennas, and M receive antennas. At time k, the time-domain symbol to be sent from the nth transmit antenna is expressed as [35]

$$\tilde{\mathbf{s}}_n(k)=\mathbf{G}\mathbf{F}^{-1}\mathbf{s}_n(k),\quad 1\leqslant n\leqslant N \tag{2.1}$$

where $\mathbf{s}_n(k)\in\mathbb{C}^Q$ is the frequency-domain OFDM symbol, $\mathbf{F}\in\mathbb{C}^{Q\times Q}$ is the FFT matrix, $\mathbf{G}:=[\bar{\mathbf{I}}_G^T\quad \mathbf{I}_Q]^T\in\mathbb{C}^{(G+Q)\times Q}$, with $\bar{\mathbf{I}}_G\in\mathbb{C}^{G\times Q}$ being the last G rows of $\mathbf{I}_Q$, accounts for the insertion of a CP with length G. Let $h^{(m,n)}(\cdot)$ be the impulse response of the channel between the nth transmit antenna and the mth receive antenna; we assume without loss of generality that all the *MN* channels are of the same order L. Then the received $(G+Q)\times 1$ time-domain data vector from the mth antenna branch is [35]

$$\tilde{\mathbf{r}}_m(k)=\sum_{n=1}^{N}\tilde{\mathbf{H}}_0^{(m,n)}\tilde{\mathbf{s}}_n(k)+\sum_{n=1}^{N}\tilde{\mathbf{H}}_1^{(m,n)}\tilde{\mathbf{s}}_n(k-1)+\tilde{\mathbf{v}}_m(k) \tag{2.2}$$

where $\tilde{\mathbf{H}}_0^{(m,n)}\in\mathbb{C}^{(G+Q)\times(G+Q)}$ is lower triangular Toeplitz with $[h^{(m,n)}(0)\cdots h^{(m,n)}(L)0\cdots 0]^T$ as the first column, $\tilde{\mathbf{H}}_1^{(m,n)}\in\mathbb{C}^{(G+Q)\times(G+Q)}$ is upper triangular Toeplitz with $[0\cdots 0h^{(m,n)}(L)\cdots h^{(m,n)}(1)]$ as the first row, and $\tilde{\mathbf{v}}_m(k)\in\mathbb{C}^{(G+Q)}$ is the noise vector. To process the received data $\tilde{\mathbf{r}}_m(k)$, the G leading guard samples is first discarded; this corresponds to post-multiplying $\tilde{\mathbf{r}}_m(k)$ by the CP-removal matrix $\mathbf{M}:=[\mathbf{O}_{Q\times G}\quad \mathbf{I}_Q]^T\in\mathbb{C}^{(Q+G)\times Q}$ to get

$$\begin{aligned}\mathbf{r}_m(k)&:=\mathbf{M}^T\tilde{\mathbf{r}}_m(k)\\&=\sum_{n=1}^{N}\underbrace{\mathbf{M}^T\tilde{\mathbf{H}}_0^{(m,n)}\mathbf{G}}_{:=\mathbf{H}_0^{(m,n)}}\mathbf{F}^{-1}\mathbf{s}_n(k)\\&\quad+\sum_{n=1}^{N}\underbrace{\mathbf{M}^T\tilde{\mathbf{H}}_1^{(m,n)}\mathbf{G}}_{:=\mathbf{H}_1^{(m,n)}}\mathbf{F}^{-1}\mathbf{s}_n(k-1)\\&\quad+\mathbf{M}^T\tilde{\mathbf{v}}_m(k).\end{aligned} \tag{2.3}$$

When the length of CP is no less than the channel order, i.e., $G \geqslant L$, the received signal $\mathbf{r}_m(k)$ is free from ISI and ICI, so that $\mathbf{H}_1^{(m,n)} = \mathbf{0}_Q$ and $\mathbf{H}_0^{(m,n)} := \mathbf{H}^{(m,n)}$, which is a circulant matrix with

$$\left[h^{(m,n)}(0)0\cdots 0h^{(m,n)}(L)\cdots h^{(m,n)}(1)\right] \tag{2.4}$$

as the first row. In this paper, we focus on the case $G < L$; as such direct manipulation shows $\mathbf{H}_1^{(m,n)}$ becomes upper triangular Toeplitz with $[0\cdots 0h^{(m,n)}(L)\cdots h^{(m,n)}(G+1)]$ as the first row, and $\mathbf{H}_0^{(m,n)}$ is instead obtained from $\mathbf{H}^{(m,n)}$ by setting $[\mathbf{H}^{(m,n)}]_{i,j} = 0$ for $1 \leqslant i \leqslant L-G$ and $Q-L+i \leqslant j \leqslant Q-G$. With (2.3), the received signal in the frequency domain reads

$$\begin{aligned}\mathbf{z}_m(k) := \mathbf{F}\mathbf{r}_m(k) &= \sum_{n=1}^{N} \mathbf{F}\mathbf{H}_0^{(m,n)}\mathbf{F}^{-1}\mathbf{s}_n(k) \\ &+ \sum_{n=1}^{N} \mathbf{F}\mathbf{H}_1^{(m,n)}\mathbf{F}^{-1}\mathbf{s}_n(k-1) \\ &+ \underbrace{\mathbf{F}\mathbf{M}^T\tilde{\mathbf{v}}_m(k)}_{:=\mathbf{v}_m(k)}.\end{aligned} \tag{2.5}$$

We note that, on the right-hand side (RHS) of (2.5), the second term is the ISI, whereas the first term is a mixture of the desired tone-by-tone signals and ICI. To realize low-complexity per-tone based signal recovery, which is one main advantage of MIMO-OFDM [21], a natural approach is to treat ISI and ICI as an overall composite interference and devise efficient schemes for *joint* ISI-ICI suppression [6], [31]. For this, it requires to further split ICI from the signal-ICI mixture in (2.5); since ICI is characterized via the deviation of $\mathbf{H}_0^{(m,n)}$ from the circulant matrix $\mathbf{H}^{(m,n)}$ [6], the splitting can be done according to the following decomposition:

$$\mathbf{H}_0^{(m,n)} = \mathbf{H}^{(m,n)} - \underbrace{\mathbf{H}_1^{(m,n)}\mathbf{J}^G}_{:=\mathbf{H}_2^{(m,n)}} \tag{2.6}$$

where

$$\mathbf{H}_2^{(m,n)} := \mathbf{H}_1^{(m,n)}\mathbf{J}^G \tag{2.7}$$

with $\mathbf{J} \in \mathbb{C}^{Q\times Q}$ denoting the circulant permutation matrix with $[0\cdots 01] \in \mathbb{C}^{1\times Q}$ as the first row [see Fig. 1 for schematic descriptions of (2.6) and (2.7)]. From (2.6), we can rewrite $\mathbf{z}_m(k)$ in (2.5) as

$$\begin{aligned}\mathbf{z}_m(k) &= \sum_{n=1}^{N} \underbrace{\mathbf{F}\mathbf{H}^{(m,n)}\mathbf{F}^{-1}}_{:=\mathbf{D}^{(m,n)}}\mathbf{s}_n(k) \\ &+ \sum_{n=1}^{N} \mathbf{F}\mathbf{H}_1^{(m,n)}\mathbf{F}^{-1}\mathbf{s}_n(k-1) \\ &- \sum_{n=1}^{N} \mathbf{F}\mathbf{H}_2^{(m,n)}\mathbf{F}^{-1}\mathbf{s}_n(k) + \mathbf{v}_m(k)\end{aligned} \tag{2.8}$$

where $\mathbf{D}^{(m,n)} \in \mathbb{C}^{Q\times Q}$ is diagonal with $[\mathbf{D}^{(m,n)}]_{ii} = \sum_{l=0}^{L} h^{(m,n)}(l)e^{-j2\pi l(i-1)/Q}$. The first term on the RHS of (2.8), which is composed of parallel tone-by-tone symbol streams from all transmit antennas, serves as the signal of interest. Since $G < L$, the symbol in each tone is contaminated by both ISI from the previously transmitted block [the second term in (2.8)] as well as ICI due to the loss of channel cyclicity [the third term in (2.8)]. Based on (2.8),[1] we propose a method for jointly suppressing ISI and ICI in the presence of channel mismatch. The following assumptions are made in the sequel.

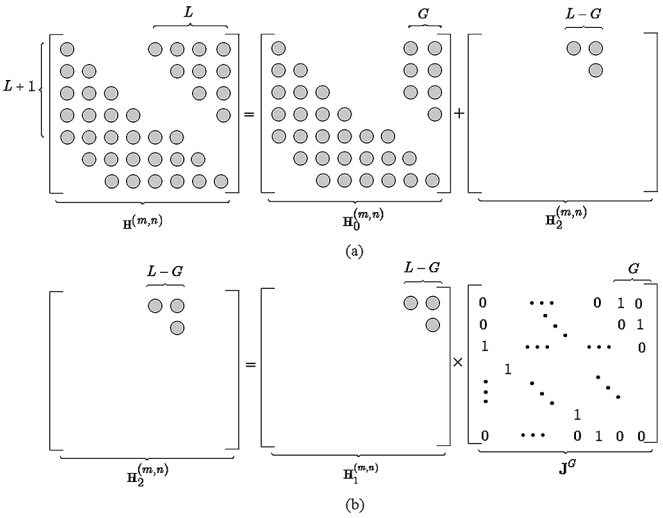

Fig. 1. Schematic descriptions of (a) the decomposition (2.6) and (b) the relation (2.7).

1) The number of receive antennas is greater than the number of transmit antennas, i.e., $M > N$.
2) The source symbols of each transmit antenna $s_n(k)$ is zero mean, unit-variance, and $E\{s_{n_1}(k_1)s_{n_2}(k_2)^*\} = \delta(n_1 - n_2)\delta(k_1 - k_2)$, where $\delta(\cdot)$ is the Kronecker delta.
3) The elements of $\tilde{\mathbf{v}}_m(k)$'s are i.i.d. complex circular Gaussian with zero mean and variance σ_v^2.

B. Least-Squares Channel Estimation

We assume that $G \geqslant L$ during the training phase and the channels are estimated based on the LS training technique; see [2], [19] for detailed treatments. Let $\mathbf{h}^{(m,n)} := [h^{(m,n)}(0)\cdots h^{(m,n)}(L)]^T \in \mathbb{C}^{L+1}$, and $\hat{\mathbf{h}}^{(m,n)} := \mathbf{h}^{(m,n)} + \Delta\mathbf{h}^{(m,n)}$ be the corresponding optimal LS estimate, with $\Delta\mathbf{h}^{(m,n)}$ modeling the estimation error. With $\Delta\mathbf{h}^{(m)} := [\Delta\mathbf{h}^{(m,1)^T}\cdots\Delta\mathbf{h}^{(m,N)^T}]^T \in \mathbb{C}^{N(L+1)}, 1 \leqslant m \leqslant M$, it is known that [2]

$$\mathbf{R}_{\mathbf{h}}^{(m)} := E\left\{\Delta\mathbf{h}^{(m)}\Delta\mathbf{h}^{(m)^H}\right\} = \frac{\sigma_v^2}{P}\mathbf{I}_{N(L+1)} \tag{2.9}$$

where P is the transmit power dedicated to channel estimation; also, since the noises between different receive antennas are independent, we have

$$\mathbf{R}_{\mathbf{h}}^{(m_1,m_2)} := E\left\{\Delta\mathbf{h}^{(m_1)}\Delta\mathbf{h}^{(m_2)^H}\right\} = \mathbf{0}_{N(L+1)}, \quad \text{for } m_1 \neq m_2. \tag{2.10}$$

The channel error properties (2.9) and (2.10) will be used in our robust equalizer design.

[1]The proposed method can also instead rely on the raw signal model (2.5) for ISI suppression first, followed by a space-frequency detector for removing ICI. Such an approach, however, can yield rather limited performance gain at the cost of significant extra computational complexity; this is further discussed in Section VII.

III. ISI-ICI MITIGATION

This section shows how the receive diversity can be exploited for ISI-ICI suppression in the considered scenario. We will first collect all the M received signal branches to form a multichannel system model, with which the interference subspace can be characterized; it will be seen that under quite mild conditions there will be sufficient inherent spatial and frequency degrees-of-freedom for combating ISI and ICI. Then we will show a constrained optimization based ISI-ICI suppression scheme under perfect channel knowledge assumption. The problem is formulated and solved within an equivalent unconstrained GSC setting, which lays the foundation of the proposed robust equalizer design.

A. Multichannel System Representation

Let $\mathbf{s}(k) := [\mathbf{s}_1^T(k) \cdots \mathbf{s}_N^T(k)]^T \in \mathbb{C}^{NQ}$ be the vector containing all the N transmitted OFDM symbol blocks. By stacking $\mathbf{z}_m(k), 1 \leqslant m \leqslant M$, in (2.8) into a vector, we can form the MQ-dimensional multichannel model:

$$\begin{aligned}\mathbf{z}(k) &= \left[\mathbf{z}_1^T(k) \cdots \mathbf{z}_M^T(k)\right]^T \\ &= \mathbf{D}\mathbf{s}(k) + \underbrace{\mathbf{F}_M \mathbf{H}_1 \mathbf{F}_N^{-1}}_{:=\mathbf{H}_{\mathrm{ISI}}} \mathbf{s}(k-1) \\ &\quad - \underbrace{\mathbf{F}_M \mathbf{H}_2 \mathbf{F}_N^{-1}}_{:=\mathbf{H}_{\mathrm{ICI}}} \mathbf{s}(k) + \mathbf{v}(k) \end{aligned} \tag{3.1}$$

where

$$\mathbf{H}_i := \left[\mathbf{H}_i^{(1)} \cdots \mathbf{H}_i^{(N)}\right] \in \mathbb{C}^{MQ \times NQ}, \quad 1 \leqslant i \leqslant 2 \tag{3.2}$$

with

$$\mathbf{H}_i^{(n)} := \left[\mathbf{H}_i^{(1,n)T} \cdots \mathbf{H}_i^{(M,n)T}\right]^T \in \mathbb{C}^{MQ \times Q}, \quad 1 \leqslant n \leqslant N \tag{3.3}$$

$$\begin{aligned}\mathbf{D} &:= [\mathbf{D}^{(1)} \cdots \mathbf{D}^{(N)}] \in \mathbb{C}^{MQ \times NQ}, \quad \text{with} \quad \mathbf{D}^{(n)} \\ &:= [\mathbf{D}^{(1,n)T} \cdots \mathbf{D}^{(M,n)T}]^T \in \mathbb{C}^{MQ \times Q}, \\ &\qquad 1 \leqslant n \leqslant N. \end{aligned} \tag{3.4}$$

$\mathbf{F}_p := \mathbf{I}_p \otimes \mathbf{F} \in \mathbb{C}^{pQ \times pQ}$, for $p \in \{M, N\}$, and $\mathbf{v}(k) := [\mathbf{v}_1^T(k) \cdots \mathbf{v}_M^T(k)]^T \in \mathbb{C}^{MQ}$. Since $M > N$, we may assume without loss of generality that the channel tone matrix $\mathbf{D}$ is of full column rank NQ; this assumption is valid whenever the MN subchannels between all transmit and receive antennas are uncorrelated. Since $\mathbf{H}_1^{(m,n)}$ is upper triangular with all the nonzero entries clustering in the last $L-G$ columns, the matrix $\mathbf{H}_1^{(n)}$ in (3.3) will have the last $L-G$ columns nonzero; by definition, the "big" $\mathbf{H}_1$ in (3.2) then has only $N(L-G)$ nonzero columns, and is of rank at most $N(L-G)$. From (2.7) and (3.3), it is easy to check $\mathbf{H}_2^{(n)} = \mathbf{H}_1^{(n)}\mathbf{J}^G$, and hence we have from (3.2) that

$$\begin{aligned}\mathbf{H}_2 &= \left[\mathbf{H}_2^{(1)}\mathbf{H}_2^{(2)} \cdots \mathbf{H}_2^{(N)}\right] \\ &= \left[\mathbf{H}_1^{(1)}\mathbf{J}^G\, \mathbf{H}_1^{(2)}\mathbf{J}^G \cdots \mathbf{H}_1^{(N)}\mathbf{J}^G\right] \\ &= \mathbf{H}_1(\mathbf{I}_N \otimes \mathbf{J}^G) \end{aligned} \tag{3.5}$$

where the last equality follows by the definition of Kronecker product. This asserts that the rank of $\mathbf{H}_2$ does not exceed $N(L-G)$ either since it is obtained by permuting the columns of $\mathbf{H}_1$. The relation (3.5) also implies the interference subspace, spanned by the columns of $\mathbf{H}_{\mathrm{ISI}}$ and $\mathbf{H}_{\mathrm{ICI}}$, is of dimension no larger than $N(L-G)$. Indeed, from (3.5) and (3.1) it follows

$$\mathbf{H}_{\mathrm{ICI}} = \mathbf{F}_M \mathbf{H}_2 \mathbf{F}_N^{-1} = \mathbf{F}_M \mathbf{H}_1 (\mathbf{I}_N \otimes \mathbf{J}^G) \mathbf{F}_N^{-1}. \tag{3.6}$$

With (3.6) and since $\mathbf{H}_{\mathrm{ISI}} = \mathbf{F}_M \mathbf{H}_1 \mathbf{F}_N^{-1}$ [see (3.1)], the column spaces of both $\mathbf{H}_{\mathrm{ISI}}$ and $\mathbf{H}_{\mathrm{ICI}}$ coincide with that of $\mathbf{F}_M\mathbf{H}_1$, whose rank is at most $N(L-G)$ (as $\mathbf{F}_M$ is orthonormal). Hence, if 1) the NQ-dimensional range space of the channel tone matrix $\mathbf{D}$ does not overlap with the ISI-ICI subspace, and 2) $(M-N)Q \geqslant N(L-G)$, it is plausible to exploit the extra degrees-of-freedom provided by the multi-channel space-frequency model (3.1) for ISI-ICI suppression. We note that condition 1) can be verified to hold unless all the MN subchannels lapse into the same direction, viz., $\mathbf{h}^{(m,n)} = \beta^{(m,n)}\mathbf{h}$ for some $\mathbf{h} \in \mathbb{C}^{L+1}$ and $\beta^{(m,n)} \in \mathbb{C}$; condition 2) is typically true since the number of subcarriers Q is often substantially larger than the channel order L.

B. GSC-Based Interference Suppression: Perfect Channel Knowledge Case

We assume for the moment that the channel is perfectly known at the receiver. To exploit the extra degrees-of-freedom in model (3.1) for interference suppression, a commonly used approach is via constrained optimization [7], [25], [30]. Specifically, we will seek for a linear weighting matrix $\mathbf{W} \in \mathbb{C}^{MQ \times NQ}$ which satisfies

$$\mathbf{W}^H\mathbf{D} = \mathbf{D}^H\mathbf{D} \tag{3.7}$$

and minimizes the mean power of the filtered ISI-ICI, i.e., $E\{\|\mathbf{W}^H(\mathbf{H}_{\mathrm{ISI}}\mathbf{s}(k-1) - \mathbf{H}_{\mathrm{ICI}}\mathbf{s}(k) + \mathbf{v}(k))\|^2\}$. With constraint (3.7), the optimal weight $\mathbf{W}$ will linearly combine the desired signal in the maximal-ratio sense (channel matched filtering), and suppress ISI-ICI via minimizes the total output interference-plus-noise power; the equalized signal then approximates

$$\mathbf{W}^H\mathbf{z}(k) \approx \mathbf{D}^H\mathbf{D}\mathbf{s}(k) \tag{3.8}$$

which can facilitate low-complexity tone-by-tone signal separation. To solve for the optimal $\mathbf{W}$, an efficient approach is to transform the constrained optimization problem into an unconstrained one via the GSC principle [8], [32]. This amounts to decomposing the weighting matrix into

$$\mathbf{W} = \mathbf{D} - \mathbf{B}\mathbf{U} \tag{3.9}$$

where the signal signature $\mathbf{D}$ represents the *nonadaptive* component for verifying constraint (3.7), $\mathbf{B} \in \mathbb{C}^{MQ \times (M-N)Q}$ is the signal blocking matrix with $\mathbf{B}^H\mathbf{D} = \mathbf{0}$, and $\mathbf{U} \in \mathbb{C}^{(M-N)Q \times NQ}$ forms the remaining free parameters

to be determined.[2] With (3.9) and (3.1), the equalized output becomes

$$\mathbf{W}^H\mathbf{z}(k) = \mathbf{z}_d(k) - \mathbf{U}^H\mathbf{z}_b(k) \tag{3.10}$$

where

$$\begin{aligned}\mathbf{z}_d(k) &:= \mathbf{D}^H\mathbf{z}(k) = \mathbf{D}^H\mathbf{D}\mathbf{s}(k) + \mathbf{D}^H\mathbf{H}_{\mathrm{ISI}}\mathbf{s}(k-1)\\ &\quad - \mathbf{D}^H\mathbf{H}_{\mathrm{ICI}}\mathbf{s}(k) + \mathbf{D}^H\mathbf{v}(k)\end{aligned} \tag{3.11}$$

and

$$\begin{aligned}\mathbf{z}_b(k) &:= \mathbf{B}^H\mathbf{z}(k) = \mathbf{B}^H\mathbf{H}_{\mathrm{ISI}}\mathbf{s}(k-1)\\ &\quad - \mathbf{B}^H\mathbf{H}_{\mathrm{ICI}}\mathbf{s}(k) + \mathbf{B}^H\mathbf{v}(k).\end{aligned} \tag{3.12}$$

Since the matched filtered branch $\mathbf{z}_d(k)$ is contaminated by

$$\mathbf{i}(k) := \mathbf{D}^H\mathbf{H}_{\mathrm{ISI}}\mathbf{s}(k-1) - \mathbf{D}^H\mathbf{H}_{\mathrm{ICI}}\mathbf{s}(k) + \mathbf{D}^H\mathbf{v}(k) \tag{3.13}$$

to effectively suppress interference equation (3.10) suggests that the matrix $\mathbf{U}$ should be chosen to render $\mathbf{U}^H\mathbf{z}_b(k)$ as close to $\mathbf{i}(k)$ as possible; more precisely, we can determine $\mathbf{U}$ by minimizing the following cost function

$$J := E\{\|\mathbf{i}(k) - \mathbf{U}^H\mathbf{z}_b(k)\|^2\}. \tag{3.14}$$

The optimal weight, denoted by $\mathbf{U}_g$, must satisfy the linear equation

$$[\mathbf{B}^H\mathbf{R}_I\mathbf{B}]\mathbf{U}_g = \mathbf{B}^H\mathbf{R}_I\mathbf{D} \tag{3.15}$$

where

$$\mathbf{R}_I := \mathbf{H}_{\mathrm{ISI}}\mathbf{H}_{\mathrm{ISI}}^H + \mathbf{H}_{\mathrm{ICI}}\mathbf{H}_{\mathrm{ICI}}^H + \sigma_v^2\mathbf{I}_{\mathrm{MQ}} \in \mathbb{C}^{MQ\times MQ}. \tag{3.16}$$

With (3.15), we have

$$\mathbf{U}_g = (\mathbf{B}^H\mathbf{R}_I\mathbf{B})^{-1}\mathbf{B}^H\mathbf{R}_I\mathbf{D} \tag{3.17}$$

and the optimal GSC weight is thus

$$\mathbf{W}_g = \mathbf{D} - \mathbf{B}\mathbf{U}_g = \mathbf{D} - \mathbf{B}(\mathbf{B}^H\mathbf{R}_I\mathbf{B})^{-1}\mathbf{B}^H\mathbf{R}_I\mathbf{D}. \tag{3.18}$$

We note that 1) the GSC scheme solves a constrained optimization problem via a simple and elegant unconstrained formulation, 2) the decomposition (3.9) alleviates algorithm complexity since only the computation of the $(M-N)Q \times NQ$ matrix $\mathbf{U}$, other than the $MQ \times NQ$ weight $\mathbf{W}$, is required, and 3) the matrix $\mathbf{W}_g$ in (3.18) is obtained based on the crucial perfect channel knowledge assumption; when channel parameter mismatch occurs due to imperfect estimation, the performance of solution (3.18) will degrade since it does not take into account channel error mitigation.

[2]The resultant GSC based solution will coincide with the optimal one derived under the original constrained-optimization based formulation if $\mathrm{rank}(\mathbf{B}) = (M-N)Q$ [3]; this condition is fulfilled if the columns of $\mathbf{B}$ form an orthonormal basis for the left null-space associated with $\mathbf{D}$.

IV. Proposed Robust Solution

This section studies the problem of robust equalizer design against channel estimation error. We will first introduce the design formulation, and point out the challenge toward a solution. Then we will characterize the estimated blocking matrix via a perturbation analysis; this is crucial for solution derivation and subsequent performance analysis. Based on the established results and assuming the channel is estimated via the LS criterion, a closed-form optimal weighting matrix is obtained and related discussions are given.

A. Problem Formulation

Our exposure will directly rely on the GSC setup. We first observe that, when only an estimated $\hat{\mathbf{D}} \neq \mathbf{D}$ is available, exact maximal-ratio combining of the desired signal $\mathbf{D}\mathbf{s}(k)$ is impossible; the best we can do, however, is to linearly combine $\mathbf{D}\mathbf{s}(k)$ through $\hat{\mathbf{D}}$ to get the approximation $\hat{\mathbf{D}}^H\mathbf{D}\mathbf{s}(k)$. This suggests us to fix $\hat{\mathbf{D}}$ as the nonadaptive portion of the GSC solution, and decompose the weighting matrix into

$$\hat{\mathbf{W}} = \hat{\mathbf{D}} - \hat{\mathbf{B}}\mathbf{U} \tag{4.1}$$

where $\hat{\mathbf{B}}$ is the blocking matrix associated with $\hat{\mathbf{D}}$, that is, $\hat{\mathbf{B}}^H\hat{\mathbf{D}} = \mathbf{0}$, and $\mathbf{U} \in \mathbb{C}^{(M-N)Q\times NQ}$ is to be determined. With (4.1), the equalized output is instead

$$\hat{\mathbf{W}}^H\mathbf{z}(k) = \bar{\mathbf{z}}_d(k) - \mathbf{U}^H\bar{\mathbf{z}}_b(k) \tag{4.2}$$

where

$$\begin{aligned}\bar{\mathbf{z}}_d(k) &:= \hat{\mathbf{D}}^H\mathbf{z}(k) = \hat{\mathbf{D}}^H\mathbf{D}\mathbf{s}(k) + \hat{\mathbf{D}}^H\mathbf{H}_{\mathrm{ISI}}\mathbf{s}(k-1)\\ &\quad - \hat{\mathbf{D}}^H\mathbf{H}_{\mathrm{ICI}}\mathbf{s}(k) + \hat{\mathbf{D}}^H\mathbf{v}(k)\end{aligned} \tag{4.3}$$

and

$$\begin{aligned}\bar{\mathbf{z}}_b(k) &:= \hat{\mathbf{B}}^H\mathbf{z}(k) = \hat{\mathbf{B}}^H\mathbf{D}\mathbf{s}(k) + \hat{\mathbf{B}}^H\mathbf{H}_{\mathrm{ISI}}\mathbf{s}(k-1)\\ &\quad - \hat{\mathbf{B}}^H\mathbf{H}_{\mathrm{ICI}}\mathbf{s}(k) + \hat{\mathbf{B}}^H\mathbf{v}(k).\end{aligned} \tag{4.4}$$

Due to inexact channel knowledge, the desired signal in the matched filtered branch $\bar{\mathbf{z}}_d(k)$ is noncoherently combined, and the contaminating interference becomes

$$\bar{\mathbf{i}}(k) := \hat{\mathbf{D}}^H\mathbf{H}_{\mathrm{ISI}}\mathbf{s}(k-1) - \hat{\mathbf{D}}^H\mathbf{H}_{\mathrm{ICI}}\mathbf{s}(k) + \hat{\mathbf{D}}^H\mathbf{v}(k). \tag{4.5}$$

The channel estimation error also modifies the blocked signal characteristics in (4.4). In particular, since the estimated $\hat{\mathbf{B}}$ is otherwise determined via $\hat{\mathbf{B}}^H\hat{\mathbf{D}} = \mathbf{0}$ (and hence $\hat{\mathbf{B}}^H\mathbf{D} \neq \mathbf{0}$ in general), there is a signal leakage $\hat{\mathbf{B}}^H\mathbf{D}\mathbf{s}(k)$ into the blocking branch $\bar{\mathbf{z}}_b(k)$. To mitigate the aggregate impacts due to channel errors, a natural approach is to treat $\bar{\mathbf{z}}_b(k)$ as a composite interference, and to design $\mathbf{U}$ by minimizing

$$\bar{J} := E\{\|\bar{\mathbf{i}}(k) - \mathbf{U}^H\bar{\mathbf{z}}_b(k)\|^2\} \tag{4.6}$$

where the expectation is taken with respect to the source signal, channel estimation error, and background noise (assuming all are mutually independent). Based on (4.4), (4.5), and (4.6), and

by averaging the cost function $\bar{J}$ over the source signal and noise, we have

$$\begin{aligned}\bar{J} = Tr\Big(\mathbf{U}^H E\Big\{\hat{\mathbf{B}}^H\Big(\mathbf{R}_I - \mathbf{D}\mathbf{H}_{\mathrm{ICI}}^H \\ - \mathbf{H}_{\mathrm{ICI}}\mathbf{D}^H + \mathbf{D}\mathbf{D}^H\Big)\hat{\mathbf{B}}\Big\}\mathbf{U}\Big) \\ - Tr\Big(\mathbf{U}^H E\Big\{\hat{\mathbf{B}}^H\Big(\mathbf{R}_I - \mathbf{D}\mathbf{H}_{\mathrm{ICI}}^H\Big)\hat{\mathbf{D}}\Big\}\Big) \\ - Tr(E\{(\hat{\mathbf{D}}^H(\mathbf{R}_I - \mathbf{H}_{\mathrm{ICI}}\mathbf{D}^H)\hat{\mathbf{B}}\}\mathbf{U}) \\ + Tr(E\{(\hat{\mathbf{D}}^H\mathbf{R}_I\hat{\mathbf{D}}\}). \end{aligned} \tag{4.7}$$

Since, for a given $\mathbf{A}$ we have[3] $\partial Tr(\mathbf{U}^H\mathbf{A}\mathbf{U})/\partial\mathbf{U} = (\mathbf{U}^H\mathbf{A})^T, \partial Tr(\mathbf{A}\mathbf{U})/\partial\mathbf{U} = \mathbf{A}^T$, and $\partial Tr(\mathbf{U}^H\mathbf{A})/\partial\mathbf{U} = \mathbf{0}_{(M-N)Q\times NQ}$, the derivative of $\bar{J}$ with respect to $\mathbf{U}$ is thus

$$\begin{aligned}\frac{\partial}{\partial\mathbf{U}}\bar{J} = \Big(\mathbf{U}^H E\Big\{\hat{\mathbf{B}}^H\Big(\mathbf{R}_I - \mathbf{D}\mathbf{H}_{\mathrm{ICI}}^H - \mathbf{H}_{\mathrm{ICI}}\mathbf{D}^H \\ + \mathbf{D}\mathbf{D}^H\Big)\hat{\mathbf{B}}\Big\}\Big)^T - (E\{(\hat{\mathbf{D}}^H(\mathbf{R}_I - \mathbf{H}_{\mathrm{ICI}}\mathbf{D}^H)\hat{\mathbf{B}}\})^T. \end{aligned} \tag{4.8}$$

With (4.8) and by setting $\partial\bar{J}/\partial\mathbf{U} = \mathbf{0}_{(M-N)Q\times NQ}$, the first-order necessary condition reads

$$\begin{aligned}\Big[E\{\hat{\mathbf{B}}^H\mathbf{D}\mathbf{D}^H\hat{\mathbf{B}})\} \\ + E\Big\{\hat{\mathbf{B}}^H\Big(\mathbf{R}_I - \mathbf{D}\mathbf{H}_{\mathrm{ICI}}^H - \mathbf{H}_{\mathrm{ICI}}\mathbf{D}^H\Big)\hat{\mathbf{B}}\Big\}\Big]\mathbf{U} \\ = E\Big\{\hat{\mathbf{B}}^H\Big(\mathbf{R}_I - \mathbf{D}\mathbf{H}_{\mathrm{ICI}}^H\Big)\hat{\mathbf{D}}\Big\}. \end{aligned} \tag{4.9}$$

Note that, with perfect channel assumption ($\hat{\mathbf{B}} = \mathbf{B}$ and $\hat{\mathbf{D}} = \mathbf{D}$), equation (4.9) reduces to (3.15), and the solution is simply given by (3.17). To determine the optimal $\mathbf{U}$ from (4.9), it is necessary to explicitly evaluate all the involved expectation terms. This can be done if we can establish an expression linking the estimated blocking matrix $\hat{\mathbf{B}}$ with the channel matrix perturbation $\Delta\mathbf{D}$. Since $\hat{\mathbf{B}}$ is constructed as a basis of the left null space associated with $\hat{\mathbf{D}}$ through SVD (recalling that $\hat{\mathbf{B}}^H\hat{\mathbf{D}} = \mathbf{0}$), an exact relation between $\hat{\mathbf{B}}$ and $\Delta\mathbf{D}$ appears highly intractable. In the next subsection, we will resort to the perturbation technique for developing an analytic (but approximate) expression.

B. Estimated Blocking Matrix: A Perturbation Analysis

Let us express the estimated channel tone matrix as

$$\hat{\mathbf{D}} = \mathbf{D} + \Delta\mathbf{D} \tag{4.10}$$

where $\Delta\mathbf{D} \in \mathbb{C}^{MQ\times NQ}$ models the estimation error and is defined similarly as $\mathbf{D}$ in (3.4), except that the (m,n)th $Q\times Q$ submatrix is

$$\Delta\mathbf{D}^{(m,n)} := \mathrm{diag}\Big\{\sqrt{Q}\mathbf{F}_L\Delta\mathbf{h}^{(m,n)}\Big\} \tag{4.11}$$

where $\mathbf{F}_L \in \mathbb{C}^{Q\times(L+1)}$ contains the first $L+1$ columns of the FFT matrix. Write an SVD of the exact signal matrix $\mathbf{D}$ as

$$\mathbf{D} = [\mathbf{U}_D \quad \mathbf{U}_B]\begin{bmatrix}\Sigma_D & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{V}_D^H\\ \mathbf{V}_B^H\end{bmatrix} \tag{4.12}$$

[3]These equalities follow immediately from the definition of the derivative of a real valued function with respect to a matrix, e.g., [34, App. A], together with some straightforward manipulations.

and likewise for $\hat{\mathbf{D}}$

$$\hat{\mathbf{D}} = [\hat{\mathbf{U}}_D\,\hat{\mathbf{U}}_B]\begin{bmatrix}\hat{\Sigma}_D & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\hat{\mathbf{V}}_D^H\\ \hat{\mathbf{V}}_B^H\end{bmatrix}. \tag{4.13}$$

We note that each component matrix in (4.13) is of the same dimension as the corresponding noise-free counterpart in (4.12); the blocking matrices with, and without, channel parameter error are, respectively, $\hat{\mathbf{B}} = \hat{\mathbf{U}}_B$ and $\mathbf{B} = \mathbf{U}_B$, both of dimension $MQ\times(M-N)Q$. Let us further express $\hat{\mathbf{B}}$ as

$$\hat{\mathbf{B}} = \mathbf{B} + \Delta\mathbf{B} \tag{4.14}$$

with $\Delta\mathbf{B}$ modeling the deviation. When $\|\Delta\mathbf{D}\|$ is small, we have the following linear first-order approximation of $\hat{\mathbf{B}}$.

Lemma 4.1 [28]: The perturbed blocking matrix $\hat{\mathbf{B}}$ can be approximated by

$$\hat{\mathbf{B}} \approx \mathbf{B} + \mathbf{U}_D\mathbf{P} \tag{4.15}$$

where $\mathbf{P} \in \mathbb{C}^{NQ\times(M-N)Q}$ is such that $\|\mathbf{P}\| = O(\|\Delta\mathbf{D}\|)$. □

To completely specify $\hat{\mathbf{B}}$ in the form (4.15), it remains to determine $\mathbf{P}$. This can be done by further taking into account the equality $\hat{\mathbf{B}}^H\hat{\mathbf{D}} = \mathbf{0}$, which together with (4.10) and (4.15) implies

$$\hat{\mathbf{B}}^H\hat{\mathbf{D}} = (\mathbf{B} + \mathbf{U}_D\mathbf{P})^H(\mathbf{D} + \Delta\mathbf{D}) = \mathbf{0}. \tag{4.16}$$

Since $\mathbf{B}^H\mathbf{D} = \mathbf{0}$ and $\mathbf{U}_D^H\mathbf{D} = \Sigma_D\mathbf{V}_D^H$ [see (4.12)], (4.16) can be rearranged into

$$\mathbf{V}_D\Sigma_D\mathbf{P} + \Delta\mathbf{D}^H\mathbf{U}_D\mathbf{P} = -\Delta\mathbf{D}^H\mathbf{B}. \tag{4.17}$$

To determine $\mathbf{P}$ from (4.17), we further observe that

$$\begin{aligned}\|\Delta\mathbf{D}^H\mathbf{U}_D\mathbf{P}\| \leqslant \|\Delta\mathbf{D}\|\cdot\|\mathbf{U}_D\mathbf{P}\| = \|\Delta\mathbf{D}\|\cdot\|\mathbf{P}\| \\ \propto O(\|\Delta\mathbf{D}\|^2) \end{aligned} \tag{4.18}$$

where the equality follows from the orthonormality of $\mathbf{U}_D$. Since $\|\mathbf{P}\| = O(\|\Delta\mathbf{D}\|)$, inequality (4.18) asserts that $\|\Delta\mathbf{D}^H\mathbf{U}_D\mathbf{P}\|$ is bounded from above by some quantity quadratic in $\|\Delta\mathbf{D}\|$, which is small with small $\|\Delta\mathbf{D}\|$. We may thus neglect $\Delta\mathbf{D}^H\mathbf{U}_D\mathbf{P}$ in (4.17) so that

$$\mathbf{V}_D\Sigma_D\mathbf{P} \approx -\Delta\mathbf{D}^H\mathbf{B}. \tag{4.19}$$

This technique is used in [15] for determining the first-order perturbation. With (4.19), the matrix $\mathbf{P}$ can thus be approximated by

$$\mathbf{P} = -\Sigma_D^{-1}\mathbf{V}_D^H\Delta\mathbf{D}^H\mathbf{B}. \tag{4.20}$$

This immediately implies $\Delta\mathbf{B} = \mathbf{U}_D\mathbf{P} = -\mathbf{U}_D\Sigma_D^{-1}\mathbf{V}_D^H\Delta\mathbf{D}^H\mathbf{B}$, and

$$\hat{\mathbf{B}} = \mathbf{B}\underbrace{-\mathbf{U}_D\Sigma_D^{-1}\mathbf{V}_D^H\Delta\mathbf{D}^H\mathbf{B}}_{:=\Delta\mathbf{B}}. \tag{4.21}$$

Equation (4.21) provides a closed-form expression of $\hat{\mathbf{B}}$ *linear* in the estimation error $\Delta\mathbf{D}$. The linearity nature can considerably simplify the derivation of the optimal solution and will also lead to tractable procedures of performance analysis. We note that, instead of (4.15), more accurate approximation of $\hat{\mathbf{B}}$ can be obtained by incorporating higher order components [37].

Although this can improve the solution accuracy, the resultant analysis would however become intractable.

C. Optimal Solution

Based on (4.21) and assuming that the channel is estimated in the optimal LS sense, we can explicitly determine the expectation quantities involved in (4.9); these are summarized in the next lemma (see Appendix A for a proof), and will be used for deriving an optimal weighting matrix.

Lemma 4.2: The following results hold.

1)
$$E\left\{\hat{\mathbf{B}}^H\left(\mathbf{R}_I-\mathbf{D}\mathbf{H}_{\mathrm{ICI}}^H\right)\hat{\mathbf{D}}\right\}=\mathbf{B}^H\mathbf{R}_I\mathbf{D}.$$

2)
$$E\{\hat{\mathbf{B}}^H\mathbf{D}\mathbf{D}^H\hat{\mathbf{B}}\}=\frac{N(L+1)\sigma_v^2}{P}\mathbf{I}_{(M-N)Q}.$$

3) Define the matrix
$$\mathbf{K}:=(\mathbf{D}^H\mathbf{D})^{-1}\mathbf{D}^H\left(\mathbf{R}_I-\mathbf{D}\mathbf{H}_{\mathrm{ICI}}^H-\mathbf{H}_{\mathrm{ICI}}\mathbf{D}^H\right)\times\mathbf{D}(\mathbf{D}^H\mathbf{D})^{-1}. \quad (4.22)$$
Then we have
$$E\left\{\hat{\mathbf{B}}^H\left(\mathbf{R}_I-\mathbf{D}\mathbf{H}_{\mathrm{ICI}}^H-\mathbf{H}_{\mathrm{ICI}}\mathbf{D}^H\right)\hat{\mathbf{B}}\right\}=\mathbf{B}^H(\mathbf{R}_I+\mathbf{R}_c)\mathbf{B}, \quad (4.23)$$
in which $\mathbf{R}_c\in\mathbb{C}^{MQ\times MQ}$ is block diagonal, with the mth $Q\times Q$ block diagonal submatrix $\mathbf{R}_c^{(m)}$ given by
$$\mathbf{R}_c^{(m)}=\frac{Q\sigma_v^2}{P}\mathbf{F}_L\mathbf{F}_L^H\odot\sum_{n=1}^{N}\mathbf{K}^{(n)} \quad (4.24)$$
and $\mathbf{K}^{(n)}$ denotes the nth $Q\times Q$ diagonal block of $\mathbf{K}$. □

Based on (4.9) and Lemma 4.2, the optimal $\mathbf{U}_{\mathrm{opt}}$ can be obtained as
$$\mathbf{U}_{\mathrm{opt}}=\left(\frac{N(L+1)\sigma_v^2}{P}\mathbf{I}_{(M-N)Q}+\mathbf{B}^H(\mathbf{R}_I+\mathbf{R}_c)\mathbf{B}\right)^{-1}\times\mathbf{B}^H\mathbf{R}_I\mathbf{D} \quad (4.25)$$
where $\mathbf{R}_I$ and $\mathbf{R}_c$ are, respectively, defined in (3.16) and (4.24). Note that solution (4.25) is *on average* the optimal choice for ISI-ICI suppression under (white) LS channel estimation error assumption. In practical implementation when only an estimated channel is available, the sampled-version of the robust GSC weight is thus
$$\hat{\mathbf{W}}_{\mathrm{opt}}=\hat{\mathbf{D}}-\hat{\mathbf{B}}\hat{\mathbf{U}}_{\mathrm{opt}}=\hat{\mathbf{D}}-\hat{\mathbf{B}}\left(\frac{N(L+1)\sigma_v^2}{P}\mathbf{I}_{(M-N)Q}+\hat{\mathbf{B}}^H(\hat{\mathbf{R}}_I+\hat{\mathbf{R}}_c)\hat{\mathbf{B}}\right)^{-1}\hat{\mathbf{B}}^H\hat{\mathbf{R}}_I\hat{\mathbf{D}} \quad (4.26)$$
in which $\hat{\mathbf{R}}_I$ and $\hat{\mathbf{R}}_c$ are respectively the estimates of $\mathbf{R}_I$ and $\mathbf{R}_c$.

Discussions:

1) The proposed approach explicitly incorporates the channel mismatch effect into the GSC formulation; it aims for joint mitigation of ISI-ICI and the net parameter mismatch effects induced by channel estimation errors. The optimal $\mathbf{U}_{\mathrm{opt}}$ differs from $\mathbf{U}_g$ in (3.17) in additional two terms, namely, $(P^{-1}N(L+1)\sigma_v^2)\mathbf{I}_{(M-N)Q}$ and $\mathbf{B}^H\mathbf{R}_c\mathbf{B}$, involved in matrix inversion; the former accounts for the signal leakage effect driven by the white-noise-like LS channel estimation error [cf. (2.9) and (2.10)], whereas the latter is due to the parameter perturbation in the ISI-ICI signature matrices.
2) We should note that our design formulation is not exclusive to the case with LS channel estimation (which produces a white channel estimation error); it does provide a unified framework for robust GSC filter design regardless of the adopted channel acquisition techniques. Indeed, as long as the channel error (with known covariance matrix) is independent of the source signal and noise, (4.9) remains true and the proposed approach will yield a solution of the form (4.25), except that all the involved matrices are accordingly modified based on the actual channel error characteristics.
3) Since $\mathbf{D}^H\mathbf{D}$ is the coherently combined signal signature, the magnitude of its nonzero entries would in general be substantially larger than those of $\mathbf{R}_I$ and $\mathbf{D}\mathbf{H}_{\mathrm{ICI}}^H$. This implies the entries of the matrix $\mathbf{K}$ in (4.22), and $\mathbf{B}^H\mathbf{R}_c\mathbf{B}$ as well, could be relatively small as compared with $P^{-1}N(L+1)\sigma_v^2$ (through simulations it is found that the entries of $\mathbf{B}^H\mathbf{R}_c\mathbf{B}$ are two-order less in magnitude in the medium-to-high SNR region). As a result, the achievable performance of $\mathbf{U}_{\mathrm{opt}}$ in (4.25) can remain largely intact if we ignore the term $\mathbf{B}^H\mathbf{R}_c\mathbf{B}$, which reflects the ISI-ICI signature perturbation and consider the sampled DL solution
$$\hat{\mathbf{W}}_{\mathrm{dl}}:=\hat{\mathbf{D}}-\hat{\mathbf{B}}\hat{\mathbf{U}}_{\mathrm{dl}} \quad (4.27)$$
with
$$\hat{\mathbf{U}}_{\mathrm{dl}}:=\left(\frac{N(L+1)\sigma_v^2}{P}\mathbf{I}_{(M-N)Q}+\hat{\mathbf{B}}^H\hat{\mathbf{R}}_I\hat{\mathbf{B}}\right)^{-1}\times\hat{\mathbf{B}}^H\hat{\mathbf{R}}_I\hat{\mathbf{D}}. \quad (4.28)$$
This would indicate that the signal leakage, on the other hand, is the dominant effect incurred by channel estimation errors. An intuitive reason for this is that, the leaking signal component into the blocking branch (4.4) will cause undesirable signal cancellation via the two-branch GSC interference-rejection mechanism, leading to a loss of the effective SINR.
4) Robust constrained-optimization-based beamformer design with background parameter error modeled as a white noise is addressed in [5]. The solution approach reported therein is to impose certain quadratic constraint on the beamforming weight so as to potentially keep down the white-noise amplification gain [5, pp. 1366–1367]. In light of this point, another plausible approach to robust GSC filter design in our context is thus
$$\min_{\mathbf{U}}E\left\{\left\|\mathbf{i}(k)-\mathbf{U}^H\mathbf{z}_b(k)\right\|^2\right\}\quad\text{subject to }\|\mathbf{U}\|=\delta\text{ for some }\delta>0. \quad (4.29)$$

The expectation in (4.29) is taken with respect to the source signal and measurement noise. The solution to (4.29) is known to be

$$\mathbf{U}_\gamma = \left(\gamma \mathbf{I}_{(M-N)Q} + \mathbf{B}^H \mathbf{R}_I \mathbf{B}\right)^{-1} \mathbf{B}^H \mathbf{R}_I \mathbf{D} \quad \text{for some } \gamma \geqslant 0. \tag{4.30}$$

The performance of solution (4.30) depends crucially on the selection of γ [5], [18], [30], [33]; however, there are in general no tractable rules for explicitly determining an optimal γ, even when the uncertainty level δ is known [18]. We note that the suboptimal alternative (4.28) is exactly the sampled-version of the DL solution (4.30), with γ set to be

$$\bar{\gamma} := N(L+1)\sigma_v^2/P. \tag{4.31}$$

It is thus conjectured that, under white channel error assumption, $\bar{\gamma}$ derived based on the presented perturbation analysis is the best choice with respect to the design criterion (4.29). Our simulation results (see Simulation D) tend to confirm this postulation.

V. Performance Analysis

This section investigates the SINR performance of the proposed robust filter (4.26). We will first derive an approximate average SINR expression in closed-form. Then we will show the proposed solution can yield an improved SINR gain over $\mathbf{W}_g$ in (3.18); in particular, the achievable SINR increment will be quantified, and based on which several key features regarding the optimal solution can be inferred.

A. SINR Evaluation

To evaluate the average SINR attained by (4.26), we will resort to the perturbation-based technique. Specifically, we will explicitly link the estimated solution with channel mismatch $\Delta\mathbf{D}$, and then invoke the LS channel error property for mean SINR evaluation. To facilitate the underlying analysis, we will neglect the term $\hat{\mathbf{B}}^H \hat{\mathbf{R}}_c \hat{\mathbf{B}}$ in (4.26) and consider the diagonal loading solution (4.27); through simulation tests, the derived results based on the simplified solution form are seen to well predict the actual SINR tendency attained by the optimal one (4.26).

To proceed, let

$$\hat{\mathbf{W}}_{\rm dl} = \mathbf{W}_{\rm dl} + \Delta\mathbf{W}_{\rm dl} \tag{5.1}$$

where $\mathbf{W}_{\rm dl}$ is the exact solution of $\hat{\mathbf{W}}_{\rm dl}$ [by substituting the true parameters $\mathbf{D}, \mathbf{B}$ and $\mathbf{R}_I$ in (4.27)] and $\Delta\mathbf{W}_{\rm dl}$ models the deviation. With (5.1), the output from the robust GSC filter is expressed in equation (5.2), as shown at the bottom of the page, where $\mathbf{s}_{\rm dl}(k)$ is the desired signal component and $\mathbf{i}_{\rm dl}(k)$ is the overall interference and noise. With (5.2), the average SINR is thus [12]

$$\mathrm{SINR}_{\rm dl} := \frac{E\{\|\mathbf{s}_{\rm dl}(k)\|^2\}}{E\{\|\mathbf{i}_{\rm dl}(k)\|^2\}} \tag{5.3}$$

where the expectation is taken with respect to the source signal, noise, and channel estimation errors. The signal power $E\{\|\mathbf{s}_{\rm dl}(k)\|^2\}$ can be directly computed by

$$\begin{aligned} E\{\|\mathbf{s}_{\rm dl}(k)\|^2\} &= E\left\{\left\|\mathbf{W}_{\rm dl}^H \mathbf{D}\mathbf{s}(k)\right\|^2\right\} \\ &= Tr\left(\mathbf{W}_{\rm dl}^H \mathbf{D}\mathbf{D}^H \mathbf{W}_{\rm dl}\right) = \|\mathbf{D}^H\mathbf{D}\|^2 \end{aligned} \tag{5.4}$$

where the last equality in (5.4) follows from the definition of $\mathbf{W}_{\rm dl}$ in (5.1) and $\mathbf{B}^H\mathbf{D} = \mathbf{0}$. The crucial step is to determine the composite interference power $E\{\|\mathbf{i}_{\rm dl}(k)\|^2\}$. Based on further perturbation analysis, we have (see Appendix B for detailed derivations)

$$\begin{aligned} P_{I,\rm dl} &:= E\{\|\mathbf{i}_{\rm dl}(k)\|^2\} \\ &= Tr\left(\mathbf{W}_{\rm dl}^H \mathbf{R}_I \mathbf{W}_{\rm dl}\right) \\ &\quad + Tr\left(E\left\{\Delta\mathbf{W}_{\rm dl}^H \mathbf{D}\mathbf{D}^H \Delta\mathbf{W}_{\rm dl}\right\}\right) \\ &\quad + Tr\left(E\left\{\Delta\mathbf{W}_{\rm dl}^H \mathbf{R}_{I,c} \Delta\mathbf{W}_{\rm dl}\right\}\right) \end{aligned} \tag{5.5}$$

where

$$\mathbf{R}_{I,c} := \mathbf{R}_I - \mathbf{D}\mathbf{H}_{\rm ICI}^H - \mathbf{H}_{\rm ICI}\mathbf{D}^H. \tag{5.6}$$

The expectation involved on the RHS on (5.5) is with respect to the channel errors. The quantity $Tr(\mathbf{W}_{\rm dl}^H \mathbf{R}_I \mathbf{W}_{\rm dl})$ in $P_{I,dl}$ is the filtered interference power under perfect channel knowledge; the remaining two arise due to channel mismatch and can be further computed as[4] (keeping only dominant terms)

$$\begin{aligned} &Tr\left(E\left\{\Delta\mathbf{W}_{\rm dl}^H \mathbf{D}\mathbf{D}^H \Delta\mathbf{W}_{\rm dl}\right\}\right) \\ &\quad + Tr\left(E\left\{\Delta\mathbf{W}_{\rm dl}^H \mathbf{R}_{I,c} \Delta\mathbf{W}_{\rm dl}\right\}\right) \\ &\quad = \bar{\gamma}\|\mathbf{U}_{\rm dl}\|^2 + \sum_{i=1}^{6} Tr(\mathbf{X}_i) - 2Re\{Tr(\mathbf{X}_7)\} \\ &\qquad - 2Re\{Tr(\mathbf{X}_8)\} - 2Re\{Tr(\mathbf{X}_9)\} \end{aligned} \tag{5.7}$$

where the matrices $\mathbf{X}_i, 1 \leqslant i \leqslant 9$, are provided in Table I. The average SINR attained by $\mathbf{W}_{\rm dl}$ can then be evaluated based on (5.4), (5.5), and (5.7).

[4]The derivations can be found at http://mimo.cm.nctu.edu.tw/Professor/Papers/derivations.pdf.

$$\begin{aligned} \mathbf{z}_{\rm dl}(k) := \hat{\mathbf{W}}_{\rm dl}^H \mathbf{z}(k) &= \hat{\mathbf{W}}_{\rm dl}^H \mathbf{D}\mathbf{s}(k) + \hat{\mathbf{W}}_{\rm dl}^H \mathbf{H}_{\rm ISI}\mathbf{s}(k-1) - \hat{\mathbf{W}}_{\rm dl}^H \mathbf{H}_{\rm ICI}\mathbf{s}(k) + \hat{\mathbf{W}}_{\rm dl}^H \mathbf{v}(k) \\ &= \underbrace{\mathbf{W}_{\rm dl}^H \mathbf{D}\mathbf{s}(k)}_{\mathbf{s}_{\rm dl}(k)} + \underbrace{\left(\Delta\mathbf{W}_{\rm dl}^H \mathbf{D}\mathbf{s}(k) + \hat{\mathbf{W}}_{\rm dl}^H \mathbf{H}_{\rm ISI}\mathbf{s}(k-1) - \hat{\mathbf{W}}_{\rm dl}^H \mathbf{H}_{\rm ICI}\mathbf{s}(k) + \hat{\mathbf{W}}_{\rm dl}^H \mathbf{v}(k)\right)}_{\mathbf{i}_{\rm dl}(k)} \end{aligned} \tag{5.2}$$

TABLE I
FORMULAS OF $\mathbf{X}_1 \sim \mathbf{X}_9$, WITH $\mathbf{X}_i = \mathbf{M}_i^H \mathbf{Y}_i \mathbf{M}_i, 1 \leqslant i \leqslant 7, \mathbf{X}_8 = \mathbf{N}_8^H \mathbf{Y}_8 \mathbf{M}_8$, AND $\mathbf{X}_9 = \mathbf{N}_9^H \mathbf{Y}_9 \mathbf{M}_9$

$\mathbf{M}_1$	$\mathbf{I}_{NQ}$	$\mathbf{Y}_1^{(n)}$	$\frac{Q\sigma_v^2}{P}\mathbf{F}_L^*\mathbf{F}_L^T \odot \sum_{m=1}^{M}\sum_{n=1}^{N}\mathbf{D}^{(m,n)}\mathbf{D}^{(m,n)H}$
$\mathbf{M}_2$	$\mathbf{B}\mathbf{U}_{dl}$	$\mathbf{Y}_2^{(m)}$	$\frac{Q\sigma_v^2}{P}\mathbf{F}_L\mathbf{F}_L^H \odot \sum_{n=1}^{N}\left[\mathbf{V}_D\Sigma_D^{-1}\mathbf{U}_D^H\mathbf{R}_{I,c}\mathbf{U}_D\Sigma_D^{-1}\mathbf{V}_D^H\right]^{(n)}$
$\mathbf{M}_3$	$\mathbf{I}_{NQ}$	$\mathbf{Y}_3^{(n)}$	$\frac{Q\sigma_v^2}{P}\mathbf{F}_L^*\mathbf{F}_L^T \odot \sum_{m=1}^{M}\left[\mathbf{R}_{I,c}+\mathbf{R}_I\mathbf{B}(\mathbf{B}^H\mathbf{R}_{I,dl}\mathbf{B})^{-1}(\mathbf{B}^H\mathbf{R}_{I,c}\mathbf{B})(\mathbf{B}^H\mathbf{R}_{I,dl}\mathbf{B})^{-1}\mathbf{B}^H\mathbf{R}_I\right]^{(m)}$
$\mathbf{M}_4$	$\mathbf{V}_D\Sigma_D^{-1}\mathbf{U}_D^H\mathbf{R}_I\mathbf{D}$	$\mathbf{Y}_4^{(n)}$	$\frac{Q\sigma_v^2}{P}\mathbf{F}_L^*\mathbf{F}_L^T \odot \sum_{m=1}^{M}\left[\mathbf{B}(\mathbf{B}^H\mathbf{R}_{I,dl}\mathbf{B})^{-1}(\mathbf{B}^H\mathbf{R}_{I,c}\mathbf{B})(\mathbf{B}^H\mathbf{R}_{I,dl}\mathbf{B})^{-1}\mathbf{B}^H\right]^{(m)}$
$\mathbf{M}_5$	$\mathbf{V}_D\Sigma_D^{-1}\mathbf{U}_D^H\mathbf{R}_{I,dl}\mathbf{B}\mathbf{U}_{dl}$	$\mathbf{Y}_5^{(n)}$	$\frac{Q\sigma_v^2}{P}\mathbf{F}_L^*\mathbf{F}_L^T \odot \sum_{m=1}^{M}\left[\mathbf{B}(\mathbf{B}^H\mathbf{R}_{I,dl}\mathbf{B})^{-1}(\mathbf{B}^H\mathbf{R}_{I,c}\mathbf{B})(\mathbf{B}^H\mathbf{R}_{I,dl}\mathbf{B})^{-1}\mathbf{B}^H\right]^{(m)}$
$\mathbf{M}_6$	$\mathbf{B}\mathbf{U}_{dl}$	$\mathbf{Y}_6^{(m)}$	$\frac{Q\sigma_v^2}{P}\mathbf{F}_L\mathbf{F}_L^H \odot \sum_{n=1}^{N}\left[\mathbf{V}_D\Sigma_D^{-1}\mathbf{U}_D^H\mathbf{R}_{I,dl}\mathbf{B}(\mathbf{B}^H\mathbf{R}_{I,dl}\mathbf{B})^{-1}(\mathbf{B}^H\mathbf{R}_{I,c}\mathbf{B})(\mathbf{B}^H\mathbf{R}_{I,dl}\mathbf{B})^{-1}\mathbf{B}^H\mathbf{R}_{I,dl}\mathbf{U}_D\Sigma_D^{-1}\mathbf{V}_D^H\right]^{(n)}$
$\mathbf{M}_7$	$\mathbf{B}\mathbf{U}_{dl}$	$\mathbf{Y}_7^{(m)}$	$\frac{Q\sigma_v^2}{P}\mathbf{F}_L\mathbf{F}_L^H \odot \sum_{n=1}^{N}\left[\mathbf{V}_D\Sigma_D^{-1}\mathbf{U}_D^H\mathbf{R}_{I,c}\mathbf{B}(\mathbf{B}\mathbf{R}_{I,dl}\mathbf{B})^{-1}\mathbf{B}\mathbf{R}_{I,dl}\mathbf{U}_D\Sigma_D^{-1}\mathbf{V}_D^H\right]^{(n)}$
$\mathbf{M}_8$	$\mathbf{V}_D\Sigma_D^{-1}\mathbf{U}_D^H\mathbf{R}_I\mathbf{W}_{dl}$	$\mathbf{Y}_8^{(n)}$	$-\frac{Q\sigma_v^2}{P}\mathbf{F}_L^*\mathbf{F}_L^T \odot \sum_{m=1}^{M}\left[\mathbf{R}_{I,c}\mathbf{B}(\mathbf{B}^H\mathbf{R}_{I,dl}\mathbf{B})^{-1}\mathbf{B}^H\right]^{(m)}$
$\mathbf{M}_9$	$\mathbf{V}_D\Sigma_D^{-1}\mathbf{U}_D^H\mathbf{R}_{I,dl}\mathbf{B}\mathbf{U}_{dl}$	$\mathbf{Y}_9^{(n)}$	$\frac{Q\sigma_v^2}{P}\mathbf{F}_L^*\mathbf{F}_L^T \odot \sum_{m=1}^{M}\left[\mathbf{B}(\mathbf{B}^H\mathbf{R}_{I,dl}\mathbf{B})^{-1}(\mathbf{B}^H\mathbf{R}_{I,c}\mathbf{B})(\mathbf{B}^H\mathbf{R}_{I,dl}\mathbf{B})^{-1}\mathbf{B}^H\right]^{(m)}$
$\mathbf{N}_8$	$\mathbf{I}_{NQ}$	$\mathbf{N}_9$	$\mathbf{V}_D\Sigma_D^{-1}\mathbf{U}_D^H\mathbf{R}_I\mathbf{D}$

B. Achievable Performance Advantage

Based on (5.5), the proposed scheme can be shown to yield an SINR advantage over $\mathbf{W}_g$ in (3.18). To see this, let us write $\hat{\mathbf{W}}_g = \mathbf{W}_g + \Delta\mathbf{W}_g$ as an estimate of $\mathbf{W}_g$. By splitting the filtered output $\mathbf{z}_g(k) := \hat{\mathbf{W}}_g^H\mathbf{z}(k)$ in the form (5.2), it is straightforward to check that the signal power is

$$E\left\{\left\|\mathbf{W}_g^H\mathbf{D}\mathbf{s}(k)\right\|^2\right\} = Tr\left(\mathbf{W}_g^H\mathbf{D}\mathbf{D}^H\mathbf{W}_g\right) = \|\mathbf{D}^H\mathbf{D}\|^2 \tag{5.8}$$

where the last equality follows from (3.18). Also, by going through essentially the same perturbation analysis as in Appendix B, the filtered interference power is verified to be

$$\begin{aligned} P_{I,g} := E\{\|\mathbf{i}_g(k)\|^2\} &= Tr\left(\mathbf{W}_g^H\mathbf{R}_I\mathbf{W}_g\right) \\ &+ Tr\left(E\left\{\Delta\mathbf{W}_g^H\mathbf{D}\mathbf{D}^H\Delta\mathbf{W}_g\right\}\right) \\ &+ Tr\left(E\left\{\Delta\mathbf{W}_g^H\mathbf{R}_{I,c}\Delta\mathbf{W}_g\right\}\right). \end{aligned} \tag{5.9}$$

From (5.4) and (5.8), we can see that the average signal levels sustained by $\mathbf{W}_{\rm dl}$ and $\mathbf{W}_g$ are identical; the SINR are thus completely determined by the respective interference powers $P_{I,\rm dl}$ and $P_{I,g}$. When σ_v^2 is small, it can be shown that $P_{I,\rm dl} < P_{I,g}$. More precisely, we have the following result (see Appendix C for a proof).

Theorem 5.1: For small σ_v^2

$$P_{I,g} - P_{I,\rm dl} \approx \bar{\gamma}^2 Tr\left(\mathbf{U}_g^H(\bar{\gamma}\mathbf{I}+\mathbf{B}^H\mathbf{R}_I\mathbf{B})^{-1}\mathbf{U}_g\right). \tag{5.10}$$

□

Equation (5.10) shows that, in the high SNR regime, the proposed robust solution can provide an SINR increment

$$\begin{aligned} \Delta\text{SINR} := \text{SINR}_{\rm dl} - \text{SINR}_{\rm g} &\approx \frac{\left\|\mathbf{D}^H\mathbf{D}\right\|^2}{P_{I,\rm dl}} - \frac{\left\|\mathbf{D}^H\mathbf{D}\right\|^2}{P_{I,g}} \\ &= \frac{\bar{\gamma}^2\|\mathbf{D}^H\mathbf{D}\|^2 Tr\left(\mathbf{U}_g^H(\bar{\gamma}\mathbf{I}+\mathbf{B}^H\mathbf{R}_I\mathbf{B})^{-1}\mathbf{U}_g\right)}{P_{I,\rm dl}P_{I,g}}. \end{aligned} \tag{5.11}$$

Our simulation results show that (5.11) does remain valid for a wide range of SNR. The analytic SINR increment (5.11) not only quantifies the performance advantage of the robust scheme (4.26), but can also reveal several associated intrinsic features. To see this, we note that

$$\begin{aligned} &Tr\left(\mathbf{U}_g^H(\bar{\gamma}\mathbf{I}+\mathbf{B}^H\mathbf{R}_I\mathbf{B})^{-1}\mathbf{U}_g\right) \\ &\quad\geqslant \frac{Tr\left(\mathbf{U}_g^H\mathbf{U}_g\right)}{(\bar{\gamma}+\lambda_{\max}(\mathbf{B}^H\mathbf{R}_I\mathbf{B}))} \\ &\quad= \frac{\|\mathbf{U}_g\|^2}{(\bar{\gamma}+\lambda_{\max}(\mathbf{B}^H\mathbf{R}_I\mathbf{B}))}. \end{aligned} \tag{5.12}$$

With (5.12), we can infer from (5.11) the inequality relation

$$\Delta\text{SINR} \geqslant \frac{\bar{\gamma}^2\|\mathbf{U}_g\|^2\cdot\|\mathbf{D}^H\mathbf{D}\|^2}{[\lambda_{\max}(\mathbf{B}^H\mathbf{R}_I\mathbf{B})+\bar{\gamma}]P_{I,\rm dl}P_{I,g}} \tag{5.13}$$

where $\lambda_{\max}(\mathbf{B}^H\mathbf{R}_I\mathbf{B})$ denotes the maximal eigenvalue associated with $\mathbf{B}^H\mathbf{R}_I\mathbf{B}$. The lower bound (5.13) leads to the following observations.

1) Let σ_v^2 be small and fixed. Through manipulation it can be shown that the incremental SINR lower bound (5.13) will

increase as P decreases. Hence, when P is small and incurs severe channel mismatch, the proposed robust equalizer (4.26) would yield significant performance gain over solution (3.18). As P increases, and hence the estimation accuracy improves, the performance gain would however become negligible (this is also seen in our simulation).

2) We can also see from (5.13) that, for fixed P and σ_v^2, the performance improvement would be limited when $\lambda_{\max}(\mathbf{B}^H\mathbf{R}_I\mathbf{B})$, which reflects the maximal power of ISI and ICI with perfect channel knowledge, is large. This is intuitively reasonable since, under severe ISI and ICI, the equalizer will largely aim for interference suppression, rather than combating the channel mismatch effects.

VI. Complexity Comparison

This section compares the algorithm complexity of the GSC solution (3.18) with that of the time-domain channel shortening/equalization (TEQ) approach [1] and the frequency-domain per-tone equalizer (PEQ) [13]; we note that the complexity of the proposed robust solution (4.26) is essentially the same with that of (3.18).

The computational cost of solution (3.18) is in solving for the blocking matrix via $\mathbf{B}^H\mathbf{D} = \mathbf{0}$ and inverting the $(M-N)Q \times (M-N)Q$ matrix $\mathbf{B}^H\mathbf{R}_I\mathbf{B}$. A low-complexity scheme for obtaining $\mathbf{B}$ which exploits the block diagonal structure of $\mathbf{D}$ can be found in [17]. As such, the total number of flop counts for computing the GSC solution (3.18) (in terms of the number of complex multiplications) is approximately

$$\begin{aligned}\mathrm{CM}_{\mathrm{GSC}} = &\frac{1}{3}(M-N)^3Q^3 \\ &+\left[MN^2+\left(2M+\frac{5}{2}\right)\right. \\ &\left.\times\,(M-N)^2+M^2\log_2^Q\right]Q^2 \\ &+\left(\frac{1}{6}M^2(2N(M-1)-N^2)\right. \\ &\left.-\frac{1}{6}(N^2-2MN)(M-N-1)^2+N^3\right)Q.\end{aligned} \tag{6.1}$$

It is noted that the computational burden of the matrix inversion involved in (3.18) can be further alleviated by resorting to the partial adaptivity (PA) implementation [9] (the details are referred to [16], [17]); this instead calls for inverting an $N(L-G)\times N(L-G)$ matrix, and can limit the flop cost to

$$\begin{aligned}\mathrm{CM}_{\mathrm{GSC,PA}} = &[MN^2+(2M+N(L-G)) \\ &\times(M-N)^2+M^2\log_2 Q]Q^2 \\ &+\left[\frac{1}{6}M^2(2N(M-1)-N^2)\right. \\ &-\frac{1}{6}(N^2-2MN)(M-N-1)^2 \\ &\left.+\,M(M-N)+N^3\right] \\ &\times Q+MN^2(L-G)^2\log_2^Q.\end{aligned} \tag{6.2}$$

The numbers of flop counts for TEQ [1] and PEQ [13], respectively, are obtained as

$$\begin{aligned}\mathrm{CM}_{\mathrm{TEQ}} = &N^3Q+NM(M+1)(L_t+L)L_t^2 \\ &+\frac{1}{3}N^3(G+1)^3+\frac{7}{2}N^2(G+1)^2 \\ &+\frac{1}{6}N^2(G+1)^2\end{aligned} \tag{6.3}$$

and

$$\begin{aligned}\mathrm{CM}_{\mathrm{PEQ}} = &\left\{\left[2(L_p+1)(L+2L_p+1)\right.\right. \\ &\left.+\frac{1}{3}(L+L_p+1)^2\right]N^3 \\ &\left.+\frac{5}{2}(L+L_p+1)N^2+\frac{1}{6}N\right\} \\ &\times(L+L_p+1)Q\end{aligned} \tag{6.4}$$

where L_t and L_p are, respectively, the TEQ and PEQ filter orders. From (6.1), (6.3), and (6.4), we have the following observations: 1) TEQ method calls for the least algorithm complexity among the three; 2) the complexities of the GSC and PEQ methods are comparable for moderate numbers of subcarriers Q; 3) the GSC solution entails the highest computational cost when Q is large. As we will see in the simulation section, the proposed GSC approach, although incurring more algorithm complexity, does yield significant performance improvement over the other two comparative methods, even when perfect channel parameters are used for equalizer design.

VII. Alternative Formulation

The proposed approach aims for *joint* mitigation of ISI and ICI through the GSC equalizer, and can facilitate low-complexity per-tone signal recovery after ISI-ICI suppression. This figure of merit benefits mainly from the decomposition (2.6), which splits the frequency-domain signal model into a sum of three components (namely, the ISI-ICI-free per-tone decoupled signals, along with ISI and ICI) as in (2.8); we note that a similar technique is also adopted in [6] and [31] for tackling interference in a similar high-rate scenario but for the SISO case. Without resorting to (2.6), the GSC based design can also be done instead via the original system (2.5), thus the resultant multichannel model

$$\mathbf{z}(k) = \underbrace{(\mathbf{D}-\mathbf{H}_{\mathrm{ICI}})}_{:=\overline{\mathbf{D}}}\mathbf{s}(k)+\mathbf{H}_{\mathrm{ISI}}\mathbf{s}(k-1)+\mathbf{v}(k) \tag{7.1}$$

where $\bar{\mathbf{D}} \in \mathbb{C}^{MQ\times NQ}$ is the signal signature. By following the same procedures as in Sections III-B and IV, the alternative GSC filter under perfect channel knowledge can be obtained as

$$\mathbf{W}_a = \overline{\mathbf{D}}-\overline{\mathbf{B}}(\overline{\mathbf{B}}^H\bar{\mathbf{R}}_{\mathrm{in}}\bar{\mathbf{B}})^{-1}\bar{\mathbf{B}}^H\bar{\mathbf{R}}_{\mathrm{in}}\bar{\mathbf{D}} \tag{7.2}$$

where $\bar{\mathbf{B}}^H\bar{\mathbf{D}} = \mathbf{0}$ and $\bar{\mathbf{R}}_{\mathrm{in}} = \mathbf{H}_{\mathrm{ISI}}\mathbf{H}_{\mathrm{ISI}}^H+\sigma_v^2\mathbf{I}_{\mathrm{MQ}}$; the associated sampled robust solution is instead

$$\hat{\mathbf{W}}_a = \hat{\bar{\mathbf{D}}}-\hat{\bar{\mathbf{B}}}\left(\hat{\bar{\mathbf{B}}}^H\left(\mathbf{F}_M\mathbf{D}_H\mathbf{F}_M^H+\hat{\bar{\mathbf{R}}}_{\mathrm{in}}\right)\hat{\bar{\mathbf{B}}}\right)^{-1}\hat{\bar{\mathbf{B}}}^H\hat{\bar{\mathbf{R}}}_{\mathrm{in}}\hat{\bar{\mathbf{D}}} \tag{7.3}$$

in which $\mathbf{F}_M$ is defined in (3.4) and $\mathbf{D}_H \in \mathbb{C}^{MQ\times MQ}$ is a diagonal matrix whose ith element is given by

$$\begin{cases} \frac{N(G+(i)_Q)\sigma_v^2}{P}, & i = mQ+j, 1 \leqslant j \leqslant L-G \\ \frac{N(L+1)\sigma_v^2}{P}, & i = mQ+j, L-G+1 \leqslant j \leqslant Q \end{cases} \quad \text{for } 0 \leqslant m \leqslant M-1$$

where $(\cdot)_Q$ denotes modulo- Q operation. Since (7.2) and (7.3) are obtained by treating $\mathbf{H}_{\mathrm{ISI}}\mathbf{s}(k-1)$ as the interference to be mitigated, only ISI will be suppressed through the GSC filtering; also, as $\bar{\mathbf{D}}$ is no longer block diagonal, a space-frequency signal detector is then necessary for separating the NQ coupled streams. The performance of such an approach could benefit from larger space-frequency diversity gains in the signal separation stage: if the VBLAST detector is used and inter-layer error propagation is negligible, the diversity gain per stage increases from 1 to NQ (the maximal achievable diversity gain for the per-tone based detection, however, is merely N). Our simulation results indicate that it outperforms (but only slightly) the proposed joint ISI-ICI suppression scheme. The main drawback of this alternative solution is the intensive algorithm complexity. Indeed, with the MMSE-VBLAST detector [11], [36], the tone-by-tone based separation in the proposed approach requires

$$\mathrm{CM_P} = \left(\frac{43}{12}N^4 + \frac{22}{3}N^3\right)Q \tag{7.4}$$

flops. However, in the alternative solution the number of flop counts for signal detection is approximately

$$\mathrm{CM_{SF}} = \frac{43}{12}N^4Q^4 + \frac{22}{3}N^3Q^3. \tag{7.5}$$

Even if the fast recursive implementation [4] is employed, the overall flop cost can only be reduced to be of $O(Q^3)$, which is significantly higher than that of the proposed scheme (especially when the number of subcarriers Q is large).

VIII. Simulation Resuls

This section uses several numerical examples to illustrate the performance of the proposed method. We consider a MIMO-OFDM system with $N = 2$ transmit antennas, $M = 3$ receive antennas, and $Q = 64$ subcarriers; the source symbols are drawn from the QPSK constellation. The background channel characteristics follow the standard wireless exponential decay model [20]: the channel impulse response is normalized such that $\sum_{n=1}^{N}\sum_{l=0}^{L} E\{|h^{(m,n)}(l)|^2\} = 1$. The input SNR at the mth receive antenna is defined as $1/\sigma_v^2$. We consider the quasi-static environment, in which the channels are assumed to remain constant during per coherent interval of 300 OFDM symbol periods, and can vary independently between different intervals. In each data burst, the training pilots are placed in the entire heading OFDM symbol and are designed according to [2]. The outputs of both TEQ and GSC filters are fed into an MMSE-VBLAST detector [11], [36] for further separating the multi-antenna transmitted signals on each tone. All the simulations results are averaged over 800 trials.

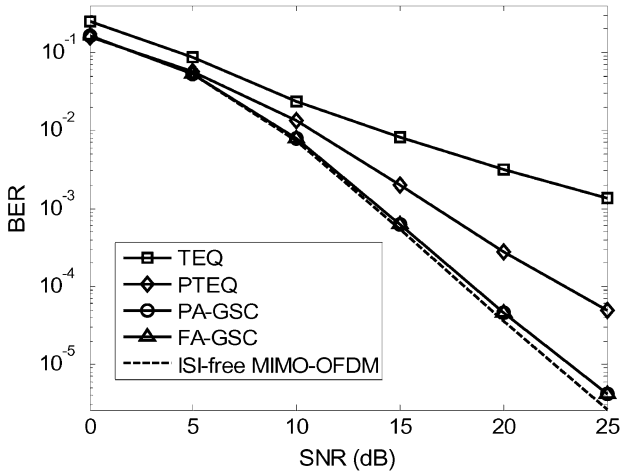

Fig. 2. BER performances of the three methods (perfect channel knowledge).

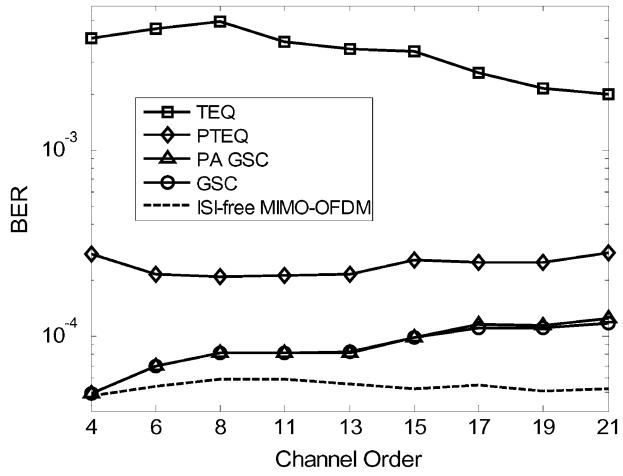

Fig. 3. BER performances of the three methods at various channel orders (perfect channel knowledge).

A. Comparison With Previous Works

We first compare the bit-error-rate (BER) performance of the proposed GSC based receiver with that of the TEQ [1] and PEQ [13] approaches. For a given channel order L, the performances of TEQ and PEQ depend crucially on the equalizer order and the allowable decision delay. The TEQ is implemented using an $(L+20)$-tap filter to shorten the composite channel order to the prescribed CP length G (through simulation it is found that further increase in the filter order does not seem to improve performance); also, the resultant delay choice yielding the lowest BER is determined through exhaustive search and is then used in simulation. The order of PEQ, and the associated decision delay, are both set to be $L-G$, as suggested in [13]. In the first simulation, we consider the case when the channel is perfectly known at the receiver [the GSC equalizer (3.18) is adopted]. For channel order $L = 13$ and CP length $G = 2$, Fig. 2 shows the BER results of the three methods at different SNR levels. We can see that, among the three ISI-ICI mitigation schemes, the GSC equalizer (3.18) yields the best performance: it incurs no

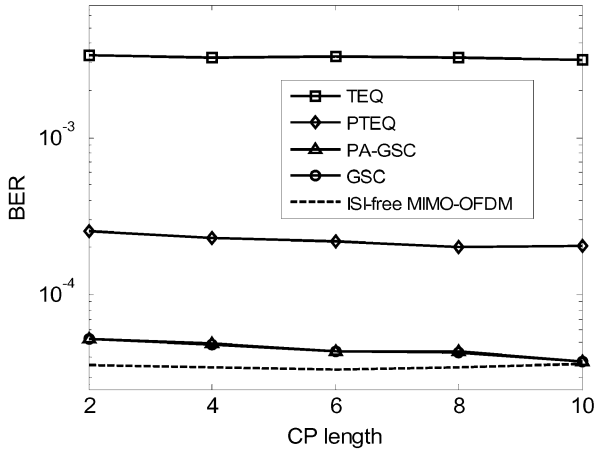

Fig. 4. BER performances of the three methods at various CP lengths (perfect channel knowledge).

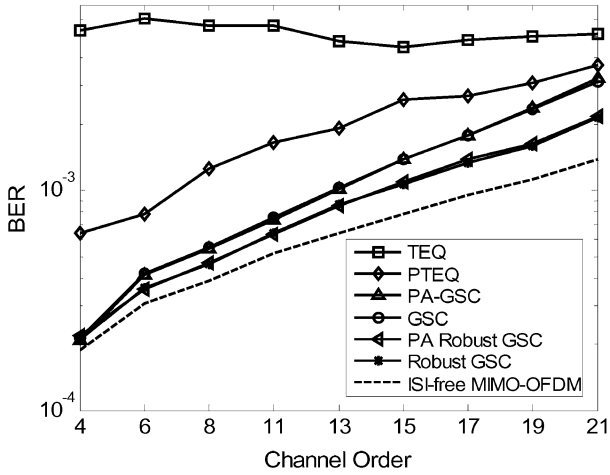

Fig. 6. BER performances of the three methods at various channel orders (LS channel estimate).

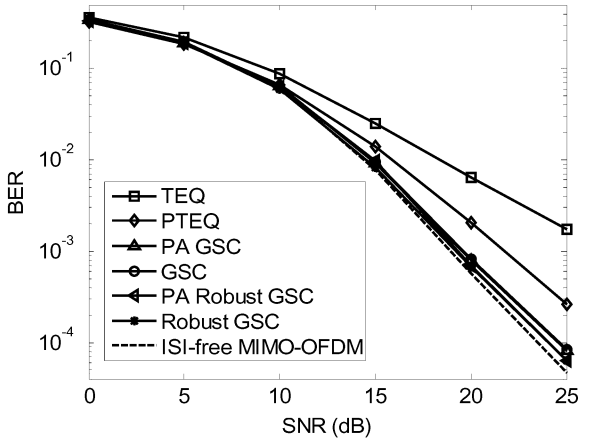

Fig. 5. BER performances of the three methods (LS channel estimate).

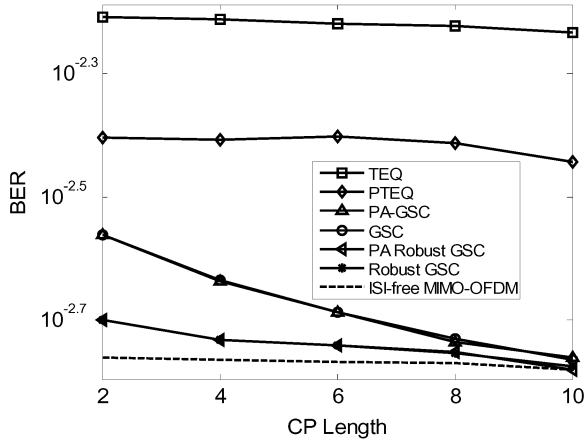

Fig. 7. BER performances of the three methods at various CP lengths (LS channel estimate).

more than 1-dB penalty as compared with the ISI-free benchmark result, and even the associated low-complexity PA implementation can outperform TEQ and PEQ. The performance advantage of the GSC approach would lie in the exploitation of the joint space-frequency degrees-of-freedom for interference suppression. With SNR $=$ 20 dB, Fig. 3 compares the respective BER as the CP length fixed at $G = 2$ and channel order L increases from 4 to 21; on the other hand, Fig. 4 shows the BER results with L fixed at 13 and CP length G varying from 2 to 10. As we can see, the GSC filter in both cases yields the lowest BER. In the second simulation, we repeat the above three experiments but instead use the LS channel estimate for equalizer design; the results are shown in Figs. 5–7 (the power dedicated for training is $P = 14$). Compared with Figs. 2–4, we can see that TEQ, PEQ, and the nonrobust GSC filter (3.18) all suffer performance degradation due to imperfect channel estimation. The proposed robust solution (4.26) is seen to improve the performance over the one (3.18); still, it maintains less than 1-dB SNR gap with respect to the ISI-free case, and can also relieve the BER penalty against long delay spread channels and short CP lengths. Finally, we demonstrate the BER performances of the three comparative methods under distinct subchannel orders: $L_{1,1} = L_{1,2} = 13, L_{2,1} = L_{2,2} = 5$, and $L_{3,1} = L_{3,2} = 11$, where $L_{m,n}$ denotes the order of the subchannel between the nth transmit and mth receive antennas, $1 \leqslant m \leqslant 3, 1 \leqslant n \leqslant 2$. In implementing the three methods we thus set $L = 13$ (for the subchannels with orders smaller than 13, $L - L_{m,n}$ zeros are padded in the respective impulse response tails). For $G = 2$, Figs. 8 and 9 show the BER performances at different SNR under perfect and imperfect channel estimates, respectively; as we can see, the BER tendencies are essentially the same with those in the common subchannel order case (cf. Figs. 2 and 5).

B. Performance of the Suboptimal Diagonal Loading Scheme (4.27)

This simulation illustrates the achievable performance of the suboptimal diagonal loading solution (4.27). For $L = 13, G = 2$, and two distinct transmit powers during the training phase $P = 14$ and 64, Fig. 10 compares solution (4.27) with the optimal weighting matrix (4.26) in terms of SINR. The results show that the respective performances are almost identical. Since the diagonal loading weight aims exclusively for signal leakage reduction, this simulated results would imply

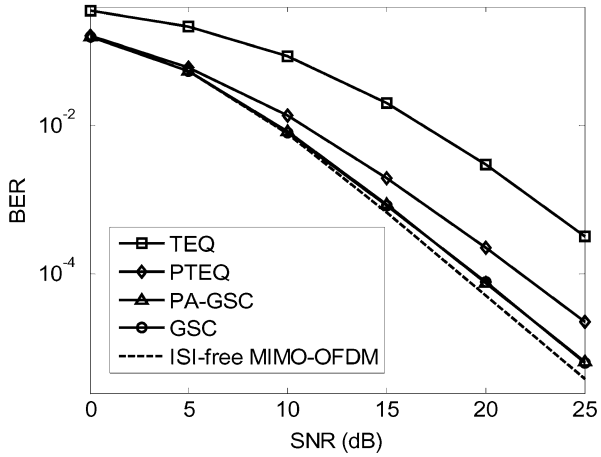

Fig. 8. BER performances of the three methods with distinct subchannel orders (perfect channel knowledge).

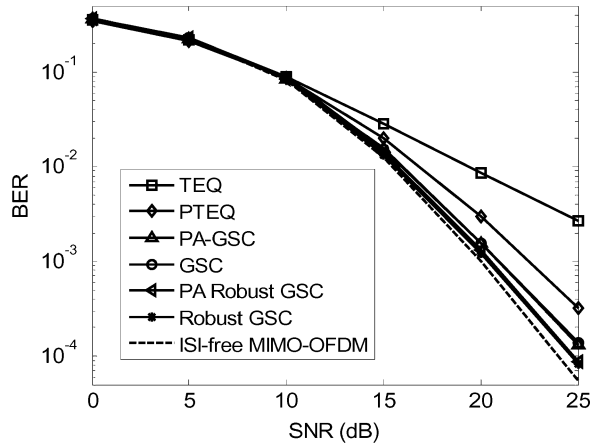

Fig. 9. BER performances of the three methods with distinct subchannel orders (LS channel estimate).

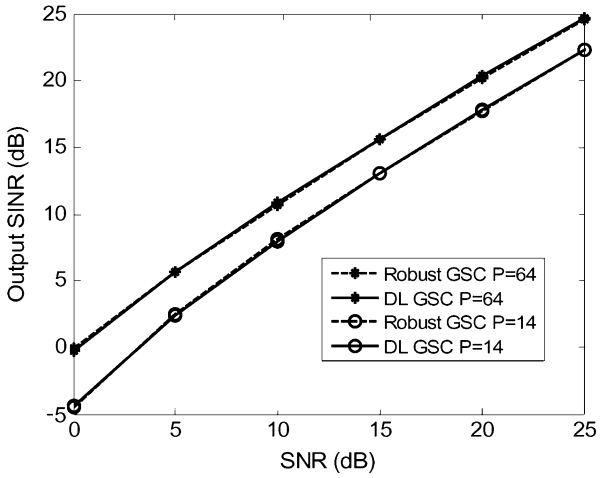

Fig. 10. SINR performances of the optimal and suboptimal DL solutions.

that the leakage effect is the prime negative factor induced by channel mismatch.

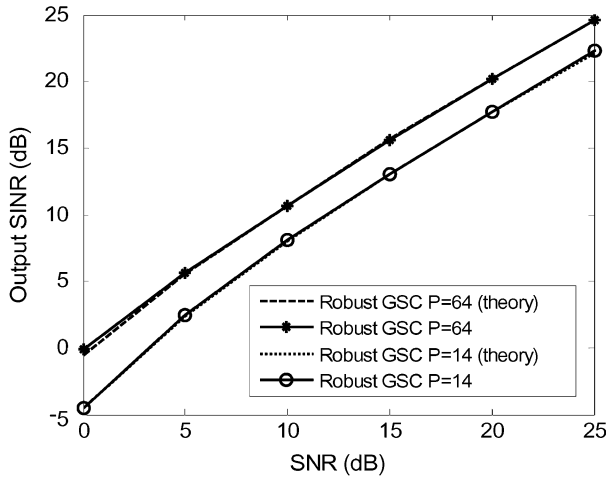

Fig. 11. Output SINR at different transmit powers in the training phase.

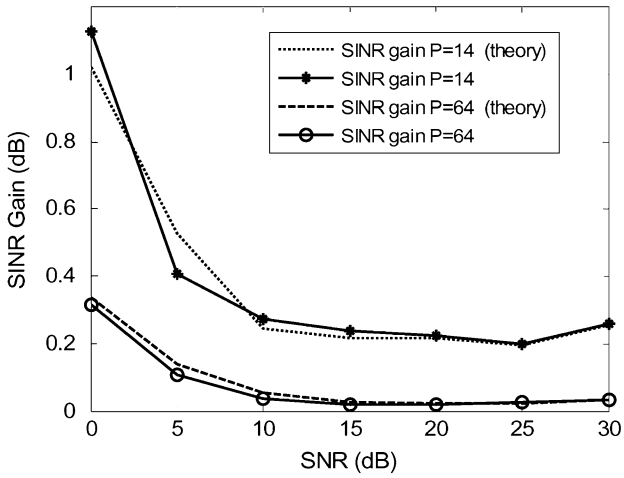

Fig. 12. SINR gain at different transmit powers in the training phase.

C. Corroboration of the Analytic SINR Result

This simulation validates the predicted SINR results in Section V. We consider two different transmit powers in the training phase $P = 14$ and $P = 64$. For $L = 13$ and $G = 2$, Fig. 11 shows the theoretical SINR, computed using the formulas in Table I, and the corresponding simulated outcomes. It can be seen that our analytic formula based on perturbation analysis well predicts the actual SINR tendency, even if the channel would be poorly estimated using a small transmit power. Fig. 12 shows the SINR gain, in terms of difference in dB, attained by the proposed robust solution (4.26) over the one (3.18); the theoretical solution is computed based on (5.11). As we can see, although the theoretical solution is derived based on the high SNR assumption, it appears very close to the experimental results over the medium-to-high SNR region (>10 dB). It is also observed that, for a fixed SNR, the SINR gain is larger for smaller P (hence, a less accurate channel estimate). This confirms the effectiveness of proposed robust GSC filter against severe parameter uncertainty; such a tendency has been deduced based on the lower bound relation (5.13) [see the first discussion following (5.13)].

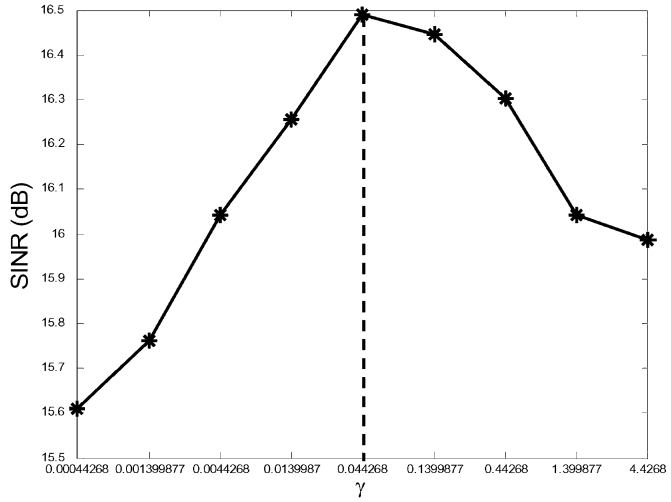

Fig. 13. Output SINR at different regularization factors $\gamma(P = 2)$.

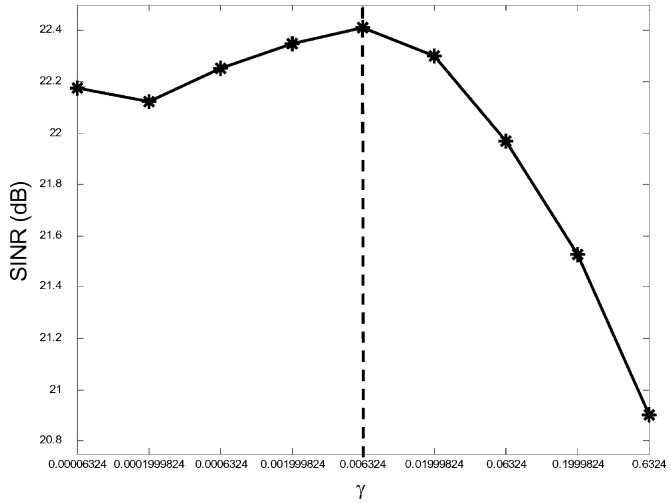

Fig. 14. Output SINR at different regularization factors $\gamma(P = 14)$.

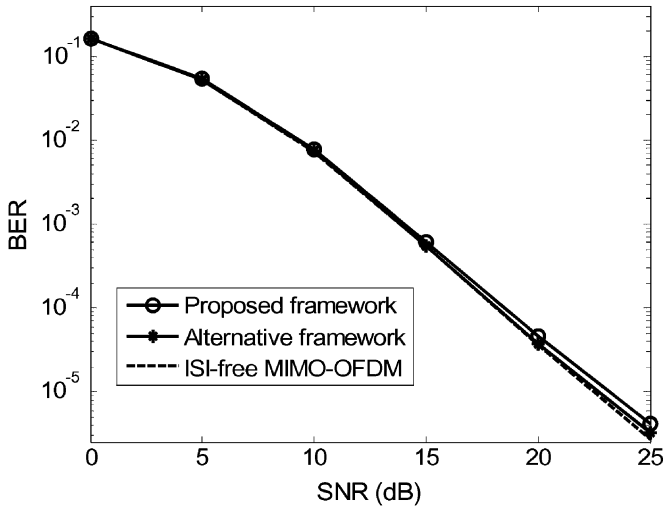

Fig. 15. BER performances of the proposed and the alternative schemes (perfect channel knowledge).

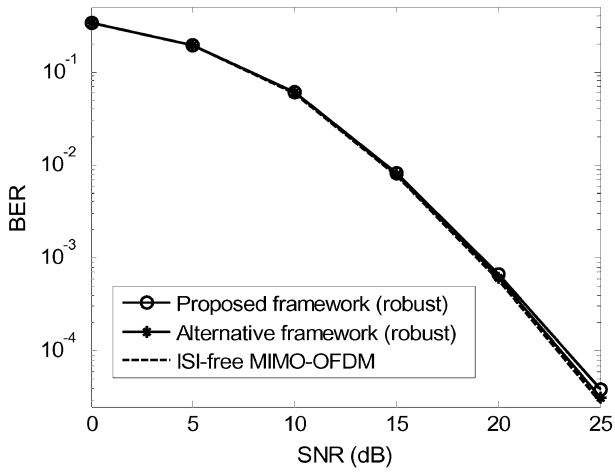

Fig. 16. BER performances of the proposed and the alternative schemes (LS channel estimate).

D. On Selection of Regularization Factor

This simulation investigates the performance of the regularization based design (4.30) at different γ factors; the channel order, CP length and noise variance are, respectively, set to be $L = 13, G = 2$ and $\sigma_v^2 = 0.003162$ (this corresponds to SNR $= 25$ dB). We consider two different transmit powers $P = 2$ and $P = 14$ in the training phase; the respective conjectured optimal $\bar{\gamma}$ computed using (4.31) are 0.044268 and 0.006324. Figs. 13 and 14 show the respective SINR performances of the regularized solution

$$\hat{\mathbf{W}}_r := \hat{\mathbf{D}} - \hat{\mathbf{B}} \left(\gamma \mathbf{I}_{(M-N)Q} + \hat{\mathbf{B}}^H \hat{\mathbf{R}}_I \hat{\mathbf{B}} \right)^{-1} \hat{\mathbf{B}}^H \hat{\mathbf{R}}_I \hat{\mathbf{D}} \tag{8.1}$$

at different values of γ. It can be seen that the SINR peaks attain at $\gamma = 0.044268$ for the $P = 2$ case, and at $\gamma = 0.006324$ when $P = 14$: this tends to indicate that $\bar{\gamma}$ in (4.31) is optimal with respect to the regularization-based design under a (stochastic) white estimation error assumption (however, $\bar{\gamma}$ will no longer be optimal when different channel error models and design criteria are considered).

E. Performance of Alternative Formulation

This simulation compares the proposed joint ISI-ICI suppression scheme with the alternative GSC equalizer discussed in Section VII in terms of BER performance; for the latter an MMSE-VBLAST detector is used for separating the NQ coupled streams (the channel order and CP length are set to be $L = 13$ and $G = 2$). Figs. 15 and 16 show the respective BER results under perfect and imperfect channel knowledge assumptions (in Fig. 16, the training power for channel estimation is $P = 14$). Although the alternative solutions (7.2) and (7.3) could benefit from a potentially larger space-frequency diversity gain in the signal detection stage, they are seen to only slightly outperform the proposed schemes (this would be due to inter-layer error propagation). We should note that the limited performance advantage attained by (7.2) and (7.3), however, comes at the price of increased computational cost (as discussed in Section VII).

IX. Conclusion

We propose a robust constrained-optimization based ISI-ICI mitigation scheme for supporting high-rate MIMO-OFDM transmission when the channels are not exactly known but are estimated via the LS training technique. The proposed constraint-free GSC design formulation yields a very natural cost function, and can facilitate the exploitation of the presumed LS channel error property toward a solution through simple first-order perturbation analysis. The proposed robust GSC filter can jointly mitigate ISI-ICI and the net detrimental factors caused by channel estimation errors. Numerical study reveals that the signal leakage is the dominant impairment and a suboptimal diagonal loading solution can attain almost all the performance gain. Based on perturbation techniques, we further derive a closed-form mean approximate SINR expression for the proposed robust scheme, and also an informative formula for quantifying the achievable SINR increment over the nonrobust solution. The analytic SINR gain reveals that prominent performance advantage can be attained by the robust solution under severe channel mismatch. Simulation results confirm the effectiveness of the proposed GSC-based equalizer: it outperforms existing methods under either perfect or imperfect channel assumption (at the cost of complexity) and, in both cases, yields a performance very close to the ISI-ICI free benchmark.

Appendix A
Proof of Lemma 4.2

Proof of (1): With $\hat{\mathbf{D}} = \mathbf{D} + \Delta\mathbf{D}$ and $\hat{\mathbf{B}}$ defined in (4.21), we have

$$\begin{aligned} & E\left\{\hat{\mathbf{B}}^H\left(\mathbf{R}_I - \mathbf{D}\mathbf{H}_{\mathrm{ICI}}^H\right)\hat{\mathbf{D}}\right\} \\ &= \mathbf{B}^H\left(\mathbf{R}_I - \mathbf{D}\mathbf{H}_{\mathrm{ICI}}^H\right)\mathbf{D} \\ &+ \mathbf{B}^H\left(\mathbf{R}_I - \mathbf{D}\mathbf{H}_{\mathrm{ICI}}^H\right)E\{\Delta\mathbf{D}\} \\ &- \mathbf{B}^H E\{\Delta\mathbf{D}\}\mathbf{V}_D\Sigma_D^{-1}\mathbf{U}_D^H \\ &\times\left(\mathbf{R}_I - \mathbf{D}\mathbf{H}_{\mathrm{ICI}}^H\right)\mathbf{D} - \mathbf{B}^H E \\ &\times\left\{\Delta\mathbf{D}\mathbf{V}_D\Sigma_D^{-1}\mathbf{U}_D^H\left(\mathbf{R}_I - \mathbf{D}\mathbf{H}_{\mathrm{ICI}}^H\right)\Delta\mathbf{D}\right\}. \end{aligned} \quad (A.1)$$

Since the elements of the noise $\mathbf{v}_m(\cdot)$ is circular Gaussian, the entries of the LS channel estimation error vector $\Delta\mathbf{h}^{(m)}$ are i.i.d. circular Gaussian [see (2.9)]. By definitions of $\Delta\mathbf{D}$ [see (4.11)] and $\Delta\mathbf{B}$ [see (4.21)] and with the circularity condition of $\Delta\mathbf{h}^{(m)}$, the last three terms on the RHS of (A.1) is identically zero. Also, since $\mathbf{v}_m(\cdot)$ is zero-mean, so are $\Delta\mathbf{h}^{(m)}$ and $\Delta\mathbf{D}$. Equation (A.1) thus reduces to

$$\begin{aligned} & E\left\{\hat{\mathbf{B}}^H\left(\mathbf{R}_I - \mathbf{D}\mathbf{H}_{\mathrm{ICI}}^H\right)\hat{\mathbf{D}}\right\} \\ &\qquad = \mathbf{B}^H\left(\mathbf{R}_I - \mathbf{D}\mathbf{H}_{\mathrm{ICI}}^H\right)\mathbf{D} = \mathbf{B}^H\mathbf{R}_I\mathbf{D} \end{aligned} \quad (A.2)$$

where the last equality follows since $\mathbf{B}^H\mathbf{D} = \mathbf{0}$.

Proof of (2): With $\hat{\mathbf{B}}$ given in (4.21), we have $E\{\hat{\mathbf{B}}^H\mathbf{D}\mathbf{D}^H\hat{\mathbf{B}}\} = \mathbf{B}^H E\{\Delta\mathbf{D}\Delta\mathbf{D}^H\}\mathbf{B}$, where the equality follows by definition of $\Delta\mathbf{B}$ in (4.21) and since $\mathbf{D} = \mathbf{U}_D\Sigma_D\mathbf{V}_D^H$. Denote by $\Delta\mathbf{D}^{(m,n)} \in \mathbb{C}^{Q\times Q}$ the (m,n)th $Q \times Q$ block submatrix of $\Delta\mathbf{D}$, for $1 \leqslant m \leqslant M$ and $1 \leqslant n \leqslant N$; then we have

$$\Delta\mathbf{D}^{(m,n)} = \mathrm{diag}\left\{\hat{\mathbf{g}}^{(m,n)}\right\}$$

where

$$\hat{\mathbf{g}}^{(m,n)} := \sqrt{Q}\mathbf{F}_L\Delta\mathbf{h}^{(m,n)} \in \mathbb{C}^Q. \quad (A.3)$$

For $\mathbf{X} \in \mathbb{C}^{m\times m}$, let $\mathrm{Diag}\{\mathbf{X}\}$ be the diagonal matrix obtained by keeping only the diagonal entries of $\mathbf{X}$. With (A.3), it follows

$$\begin{aligned} & E\left\{\Delta\mathbf{D}^{(m_1,n)}\Delta\mathbf{D}^{(m_2,n)^H}\right\} \\ &= E\left\{\mathrm{Diag}\left\{\hat{\mathbf{g}}^{(m_1,n)}\hat{\mathbf{g}}^{(m_2,n)H}\right\}\right\} \\ &= \mathrm{Diag}\left\{Q\mathbf{F}_L E\left\{\Delta\mathbf{h}^{(m_1,n)}\Delta\mathbf{h}^{(m_2,n)^H}\right\}\mathbf{F}_L^H\right\}. \end{aligned} \quad (A.4)$$

Equations (2.9) and (2.10) together imply $E\{\Delta\mathbf{h}^{(m_1,n)}\Delta\mathbf{h}^{(m_2,n)^H}\} = (P^{-1}\sigma_v^2)\mathbf{I}_{L+1}\delta(m_1 - m_2)$, and (A.4) thus becomes

$$\begin{aligned} & E\left\{\Delta\mathbf{D}^{(m_1,n)}\Delta\mathbf{D}^{(m_2,n)^H}\right\} \\ &= \frac{Q\sigma_v^2}{P}\mathrm{Diag}\left\{\mathbf{F}_L\mathbf{F}_L^H\right\}\delta(m_1 - m_2) \\ &= \frac{\sigma_v^2(L+1)}{P}\mathbf{I}_Q\delta(m_1 - m_2) \end{aligned} \quad (A.5)$$

where the last equality follows due to $\mathrm{Diag}\{\mathbf{F}_L\mathbf{F}_L^H\} = Q^{-1}(L+1)\mathbf{I}_Q$. Based on (A.5), the (p,q)th $Q \times Q$ block submatrix of $E\{\Delta\mathbf{D}\Delta\mathbf{D}^H\}$ is given by

$$\sum_{n=1}^{N} E\left\{\Delta\mathbf{D}^{(p,n)}\Delta\mathbf{D}^{(q,n)H}\right\} = \begin{cases} \frac{N(L+1)\sigma_v^2}{P}\mathbf{I}_Q, & p = q, \\ \mathbf{0}_Q, & \text{if } p \neq q. \end{cases} \quad (A.6)$$

The result follows directly from (A.6) and since $\mathbf{B}^H\mathbf{B} = \mathbf{I}_{(M-N)Q}$.

Proof of (3): With $\hat{\mathbf{B}}$ given in (4.21) and since $\mathbf{B}^H\mathbf{D} = \mathbf{0}$, direct manipulation shows

$$\begin{aligned} & E\{\hat{\mathbf{B}}^H\left(\mathbf{R}_I - \mathbf{D}\mathbf{H}_{\mathrm{ICI}}^H - \mathbf{H}_{\mathrm{ICI}}\mathbf{D}^H\right)\hat{\mathbf{B}}\} \\ &\qquad = \mathbf{B}^H\mathbf{R}_I\mathbf{B} + \mathbf{B}^H\underbrace{E\{\Delta\mathbf{D}\mathbf{K}\Delta\mathbf{D}^H\}}_{=\mathbf{R}_c}\mathbf{B} \end{aligned} \quad (A.7)$$

where the matrix $\mathbf{K}$ is defined in (4.22). We note that the (m_1,m_2)th block submatrix of $\mathbf{R}_c$ equals $\mathbf{R}_c^{(m_1,m_2)} = \sum_{n_1=1}^{N}\sum_{n_2=1}^{N} E\{\Delta\mathbf{D}^{(m_1,n_1)}\mathbf{K}^{(n_1,n_2)}\Delta\mathbf{D}^{(m_2,n_2)H}\}, 1 \leqslant m_1, m_2 \leqslant M$. Since the channel estimation errors between different receive antennas are independent [From (2.9) and

(2.10)], we have $E\{\Delta\mathbf{D}^{(m_1,n_1)}\mathbf{K}^{(n_1,n_2)}\Delta\mathbf{D}^{(m_2,n_2)H}\} = \mathbf{0}_Q, m_1 \neq m_2, n_1 \neq n_2$, which implies equation (A.8), as shown at the bottom of the page. Equation (A.8) shows $\mathbf{R}_c$ is a block diagonal matrix, and (4.24) follows by expanding the block diagonal terms in (A.8) and using $\Delta\mathbf{D}^{(m,n)} = \mathrm{diag}\{\sqrt{Q}\mathbf{F}_L\Delta\mathbf{h}^{(m,n)}\}$. □

Appendix B
Derivation of (5.5)

To derive (5.5), we need the following lemma.[5]

Lemma B.1: Let $\mathbf{R}_I$ be defined in (3.16) and $\mathbf{U}_{\mathrm{dl}}$ be the exact solution of $\hat{\mathbf{U}}_{\mathrm{dl}}$ defined in (4.28). Also, let an SVD of the channel tone matrix $\mathbf{D}$ be given in (4.12). Then the first-order approximation to $\Delta\mathbf{W}_{\mathrm{dl}}$ is

$$\begin{aligned}\Delta\mathbf{W}_{\mathrm{dl}} = {} & \Delta\mathbf{D} + \mathbf{U}_D\Sigma_D^{-1}\mathbf{V}_D^H\Delta\mathbf{D}^H\mathbf{B}\mathbf{U}_{\mathrm{dl}} \\ & - \mathbf{B}(\mathbf{B}^H\mathbf{R}_{I,\mathrm{dl}}\mathbf{B})^{-1}\left(\mathbf{B}^H\mathbf{R}_I\Delta\mathbf{D}\right. \\ & \left. - \mathbf{B}^H\Delta\mathbf{D}\mathbf{V}_D\Sigma_D^{-1}\mathbf{U}_D^H\mathbf{R}_I\mathbf{D} + \Delta\mathbf{R}\mathbf{U}_{\mathrm{dl}}\right)\end{aligned} \tag{B.1}$$

with

$$\mathbf{R}_{I,\mathrm{dl}} := \mathbf{R}_I + \bar{\gamma}\mathbf{I}_{\mathrm{MQ}} \tag{B.2}$$

and

$$\begin{aligned}\Delta\mathbf{R} := {} & \mathbf{B}^H\mathbf{R}_{I,\mathrm{dl}}\mathbf{U}_D\Sigma_D^{-1}\mathbf{V}_D^H\Delta\mathbf{D}^H\mathbf{B} \\ & + \mathbf{B}^H\Delta\mathbf{D}\mathbf{V}_D\Sigma_D^{-1}\mathbf{U}_D^H\mathbf{R}_{I,\mathrm{dl}}\mathbf{B}.\end{aligned} \tag{B.3}$$

□

Derivation of (5.5): By definitions of $\Delta\mathbf{D}$ [see (4.11)], it is easy to verify from (B.1) that $E\{\Delta\mathbf{W}_{\mathrm{dl}}\} = \mathbf{0}$; (5.5) can be obtained via substituting $\Delta\mathbf{W}_{\mathrm{dl}}$ into $\mathbf{i}_{\mathrm{dl}}(k)$ in (5.2) followed by some direct manipulations. □

Appendix C
Proof of Theorem 5.1

We shall note that the matrix $\mathbf{R}_{I,c}$ will be sparse for small σ_v^2 [cf. (5.6) and (3.16)]. This implies that, as $\Delta\mathbf{W}_{\mathrm{dl}}$ and $\Delta\mathbf{W}_g$ are small, both $Tr(E\{\Delta\mathbf{W}_{\mathrm{dl}}^H\mathbf{R}_{I,c}\Delta\mathbf{W}_{\mathrm{dl}}\})$ and $Tr(E\{\Delta\mathbf{W}_g^H\mathbf{R}_{I,c}\Delta\mathbf{W}_g\})$ are close to zero in the high SNR region, and hence

$$\begin{aligned}P_{I,g} - P_{I,\mathrm{dl}} \approx {} & Tr\left(\mathbf{W}_g^H\mathbf{R}_I\mathbf{W}_g\right) \\ & + Tr\left(E\left\{\Delta\mathbf{W}_g^H\mathbf{D}\mathbf{D}^H\Delta\mathbf{W}_g\right\}\right) \\ & - Tr\left(\mathbf{W}_{\mathrm{dl}}^H\mathbf{R}_I\mathbf{W}_{\mathrm{dl}}\right) \\ & - Tr\left(E\left\{\Delta\mathbf{W}_{\mathrm{dl}}^H\mathbf{D}\mathbf{D}^H\Delta\mathbf{W}_{\mathrm{dl}}\right\}\right).\end{aligned} \tag{C.1}$$

[5]The detailed proof can be found at http://mimo.cm.nctu.edu.tw/Professor/Papers/derivations.pdf.

With $\Delta\mathbf{W}_{\mathrm{dl}}$ in (B.1) and since $\mathbf{B}^H\mathbf{D} = \mathbf{0}$, we have $\Delta\mathbf{W}_{\mathrm{dl}}^H\mathbf{D} = \Delta\mathbf{D}^H\mathbf{D} - \mathbf{U}_{\mathrm{dl}}^H\mathbf{B}^H\Delta\mathbf{D}$, and similarly $\Delta\mathbf{W}_g^H\mathbf{D} = \Delta\mathbf{D}^H\mathbf{D} - \mathbf{U}_g^H\mathbf{B}^H\Delta\mathbf{D}$. This implies

$$\begin{aligned}& Tr\left(E\left\{\Delta\mathbf{W}_{\mathrm{dl}}^H\mathbf{D}\mathbf{D}^H\Delta\mathbf{W}_{\mathrm{dl}}\right\}\right) \\ & \quad = Tr(E\{\Delta\mathbf{D}^H\mathbf{D}\mathbf{D}^H\Delta\mathbf{D}\}) \\ & \qquad + Tr\left(\mathbf{U}_{\mathrm{dl}}^H\mathbf{B}^HE\{\Delta\mathbf{D}\Delta\mathbf{D}^H\}\mathbf{B}\mathbf{U}_{\mathrm{dl}}\right) \\ & \qquad + Tr(E\{\Delta\mathbf{D}^H\mathbf{D}\Delta\mathbf{D}^H\}\mathbf{B}\mathbf{U}_{\mathrm{dl}}) \\ & \qquad + Tr\left(\mathbf{U}_{\mathrm{dl}}^H\mathbf{B}^HE\{\Delta\mathbf{D}\mathbf{D}^H\Delta\mathbf{D}\}\right)\end{aligned} \tag{C.2}$$

and

$$\begin{aligned}& Tr\left(E\left\{\Delta\mathbf{W}_g^H\mathbf{D}\mathbf{D}^H\Delta\mathbf{W}_g\right\}\right) \\ & \quad = Tr(E\{\Delta\mathbf{D}^H\mathbf{D}\mathbf{D}^H\Delta\mathbf{D}\}) \\ & \qquad + Tr\left(\mathbf{U}_g^H\mathbf{B}^HE\{\Delta\mathbf{D}\Delta\mathbf{D}^H\}\mathbf{B}\mathbf{U}_g\right) \\ & \qquad + Tr(E\{\Delta\mathbf{D}^H\mathbf{D}\Delta\mathbf{D}^H\}\mathbf{B}\mathbf{U}_g) \\ & \qquad + Tr\left(\mathbf{U}_g^H\mathbf{B}^HE\{\Delta\mathbf{D}\mathbf{D}^H\Delta\mathbf{D}\}\right).\end{aligned} \tag{C.3}$$

The circularity condition of $\Delta\mathbf{h}^{(m)}$ implies the last two terms on the RHS of both (C.2) and (C.3) are identically zero. From (C.2), (C.3), and since $\mathbf{B}^HE\{\Delta\mathbf{D}\Delta\mathbf{D}^H\}\mathbf{B} = \bar{\gamma}$ [see (A.3) and (A.6)], we have

$$\begin{aligned}& Tr\left(E\left\{\Delta\mathbf{W}_g^H\mathbf{D}\mathbf{D}^H\Delta\mathbf{W}_g\right\}\right) \\ & \quad - Tr\left(E\left\{\Delta\mathbf{W}_{\mathrm{dl}}^H\mathbf{D}\mathbf{D}^H\Delta\mathbf{W}_{\mathrm{dl}}\right\}\right) \\ & \quad = \bar{\gamma}\left\{Tr\left(\mathbf{U}_g^H\mathbf{U}_g\right) - Tr\left(\mathbf{U}_{\mathrm{dl}}^H\mathbf{U}_{\mathrm{dl}}\right)\right\}.\end{aligned} \tag{C.4}$$

With (C.1) and (C.4), it follows

$$\begin{aligned}P_{I,g} - P_{I,\mathrm{dl}} \approx {} & Tr\left(\mathbf{W}_g^H\mathbf{R}_I\mathbf{W}_g\right) - Tr\left(\mathbf{W}_{\mathrm{dl}}^H\mathbf{R}_I\mathbf{W}_{\mathrm{dl}}\right) \\ & + \bar{\gamma}\left\{Tr\left(\mathbf{U}_g^H\mathbf{U}_g\right) - Tr\left(\mathbf{U}_{\mathrm{dl}}^H\mathbf{U}_{\mathrm{dl}}\right)\right\}.\end{aligned} \tag{C.5}$$

According to the matrix inversion lemma [27], we have

$$\begin{aligned}& \left(\mathbf{B}^H\mathbf{R}_I\mathbf{B} + \bar{\gamma}\mathbf{I}_{(M-N)Q}\right)^{-1} \\ & \quad = (\mathbf{B}^H\mathbf{R}_I\mathbf{B})^{-1} - (\mathbf{B}^H\mathbf{R}_I\mathbf{B})^{-1} \\ & \qquad \times \left[\bar{\gamma}^{-1}\mathbf{I}_{(M-N)Q} + (\mathbf{B}^H\mathbf{R}_I\mathbf{B})^{-1}\right]^{-1}\left(\mathbf{B}^H\mathbf{R}_I\mathbf{B}\right)^{-1}.\end{aligned} \tag{C.6}$$

Using (C.6) and by definitions of $\mathbf{U}_{\mathrm{dl}}$ and $\mathbf{U}_g$ [see Lemma B.1 and (3.17)], it can be verified that

$$\mathbf{U}_{\mathrm{dl}} = \mathbf{U}_g - \left(\mathbf{I}_{(M-N)Q} + \bar{\gamma}^{-1}\mathbf{B}^H\mathbf{R}_I\mathbf{B}\right)^{-1}\mathbf{U}_g \tag{C.7}$$

$$\mathbf{R}_c^{(m_1,m_2)} = \begin{cases}\sum_{n=1}^{N} E\left\{\Delta\mathbf{D}^{(m_1,n)}\mathbf{K}^{(n,n)}\Delta\mathbf{D}^{(m_1,n)H}\right\}, & m_1 = m_2 \\ \mathbf{0}_Q, & \text{otherwise}\end{cases} \tag{A.8}$$

and hence

$$\mathbf{W}_{\mathrm{dl}} = \mathbf{D} - \mathbf{B}\mathbf{U}_{\mathrm{dl}} = \underbrace{\mathbf{D} - \mathbf{B}\mathbf{U}_g}_{\mathbf{W}_g} + \mathbf{B}\left(\mathbf{I}_{(M-N)Q} + \bar{\gamma}^{-1}\mathbf{B}^H\mathbf{R}_I\mathbf{B}\right)^{-1}\mathbf{U}_g. \quad \text{(C.8)}$$

Substituting (C.7) and (C.8) into (C.5), we have

$$\begin{aligned} P_{I,g} - P_{I,\mathrm{dl}} &= 2\bar{\gamma}Tr\left(\mathbf{U}_g^H\left(\mathbf{I}_{(M-N)Q} + \bar{\gamma}^{-1}\mathbf{B}^H\mathbf{R}_I\mathbf{B}\right)^{-1}\mathbf{U}_g\right) \\ &\quad - \bar{\gamma}Tr\left(\mathbf{U}_g^H\left(\mathbf{I}_{(M-N)Q} + \bar{\gamma}^{-1}\mathbf{B}^H\mathbf{R}_I\mathbf{B}\right)^{-2}\mathbf{U}_g\right) \\ &\quad - Tr\left(\mathbf{U}_g^H\left(\mathbf{I}_{(M-N)Q} + \bar{\gamma}^{-1}\mathbf{B}^H\mathbf{R}_I\mathbf{B}\right)^{-1} \right. \\ &\quad \left. \times\mathbf{B}^H\mathbf{R}_I\mathbf{B}\left(\mathbf{I}_{(M-N)Q} + \bar{\gamma}^{-1}\mathbf{B}^H\mathbf{R}_I\mathbf{B}\right)^{-1}\mathbf{U}_g\right) \end{aligned} \quad \text{(C.9)}$$

We observe that the second and the third terms on the RHS of (C.9) can be further combined into

$$\begin{aligned} &- \bar{\gamma}Tr\left(\mathbf{U}_g^H\left(\mathbf{I}_{(M-N)Q} + \bar{\gamma}^{-1}\mathbf{B}^H\mathbf{R}_I\mathbf{B}\right)^{-2}\mathbf{U}_g\right) \\ &\quad - Tr\left(\mathbf{U}_g^H\left(\mathbf{I}_{(M-N)Q} + \bar{\gamma}^{-1}\mathbf{B}^H\mathbf{R}_I\mathbf{B}\right)^{-1} \right. \\ &\quad \left. \times\ \mathbf{B}^H\mathbf{R}_I\mathbf{B}\left(\mathbf{I}_{(M-N)Q} + \bar{\gamma}^{-1}\mathbf{B}^H\mathbf{R}_I\mathbf{B}\right)^{-1}\mathbf{U}_g\right) \\ &\quad = -\bar{\gamma}Tr\left(\mathbf{U}_g^H\left(\mathbf{I}_{(M-N)Q} + \bar{\gamma}^{-1}\mathbf{B}^H\mathbf{R}_I\mathbf{B}\right)^{-1}\mathbf{U}_g\right). \end{aligned} \quad \text{(C.10)}$$

With (C.9) and (C.10), it follows that

$$\begin{aligned} P_{I,g} - P_{I,\mathrm{dl}} &\approx 2\bar{\gamma}Tr\left(\mathbf{U}_g^H\left(\mathbf{I}_{(M-N)Q} + \bar{\gamma}^{-1}\mathbf{B}^H\mathbf{R}_I\mathbf{B}\right)^{-1}\mathbf{U}_g\right) \\ &\quad - \bar{\gamma}Tr\left(\mathbf{U}_g^H\left(\mathbf{I}_{(M-N)Q} + \bar{\gamma}^{-1}\mathbf{B}^H\mathbf{R}_I\mathbf{B}\right)^{-1}\mathbf{U}_g\right) \\ &= \bar{\gamma}Tr\left(\mathbf{U}_g^H\left(\mathbf{I}_{(M-N)Q} + \bar{\gamma}^{-1}\mathbf{B}^H\mathbf{R}_I\mathbf{B}\right)^{-1}\mathbf{U}_g\right) \\ &= \bar{\gamma}^2Tr\left(\mathbf{U}_g^H\left(\bar{\gamma}\mathbf{I}_{(M-N)Q} + \mathbf{B}^H\mathbf{R}_I\mathbf{B}\right)^{-1}\mathbf{U}_g\right). \end{aligned}$$

□

REFERENCES

[1] N. Al-Dhahir, "FIR channel-shortening equalizer for MIMO ISI channels," *IEEE Trans. Commun.*, vol. 49, no. 2, pp. 213–218, Feb. 2001.

[2] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO-OFDM systems in mobile wireless channels," *IEEE Trans. Signal Process.*, vol. 51, no. 6, pp. 1615–1624, Jun. 2003.

[3] B. R. Breed and J. Strauss, "A short proof of the equivalence of LCMV and GSC beamforming," *IEEE Signal Process. Lett.*, vol. 9, no. 6, pp. 168–169, Jun. 2002.

[4] J. Benesty, Y. Huang, and J. Chen, "A fast recursive algorithm for optimum sequential signal detection in a BLAST system," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1722–1730, Jul. 2003.

[5] H. Cox, R. Zeskind, and M. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 10, pp. 1365–1376, Oct. 1987.

[6] S. Chen and C. Zhu, "ICI and ISI analysis and mitigation for OFDM systems with insufficient cyclic prefix in time-varying channels," *IEEE Trans. Consum. Electron.*, vol. 50, no. 2, pp. 78–83, Feb. 2004.

[7] O. L. Frost, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.

[8] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antenna Propag.*, vol. AP-30, no. 1, pp. 27–34, Jan. 1982.

[9] S. Haykin and A. Steinhardt, *Adaptive Radar Detection and Estimation*. New York: Wiley, 1992.

[10] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1995.

[11] H. Jafarkhani, *Space-Time Coding: Theory and Practice*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[12] A. Leke and J. M. Cioffi, "Impact of imperfect channel knowledge on the performance of multicarrier systems," *Proc. IEEE GLOBECOM*, vol. 2, pp. 951–955, Nov. 1998.

[13] G. Leus and M. Moonen, "Per-tone equalization for MIMO OFDM systems," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2965–2975, Nov. 2003.

[14] J. Li, P. Stoica, and Z. Wang, "On robust Capon beamforming and diagonal loading," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1702–1715, Jul. 2003.

[15] F. Li, H. Liu, and R. J. Vaccaro, "Performance analysis for DOA estimation algorithms: Unification, simplification, and observations," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, no. 4, pp. 1170–1184, Oct. 1993.

[16] C. Y. Lin, J. Y. Wu, and T. S. Lee, "GSC-based frequency-domain equalizer for CP-free OFDM systems," in *Proc. IEEE ICC*, Seoul, Korea, May 2005, vol. 2, pp. 1132–1136.

[17] C. Y. Lin and T. S. Lee, "An efficient interference cancellation scheme for CP-free MIMO-OFDM systems," in *Proc. IEEE VTC-Fall 2005*, Dallas, TX, Sep. 2005, vol. 1, pp. 439–443.

[18] R. G. Lorenz and S. P. Boyd, "Robust minimum variance beamforming," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1684–1696, May 2005.

[19] J. H. Manton, "Optimal training sequences and pilot tones for OFDM systems," *IEEE Commun. Lett.*, vol. 5, no. 4, pp. 151–153, Apr. 2001.

[20] B. O'Hara and A. Petrick, *The IEEE 802.11 Handbook: A Designer's Companion*. New York: IEEE Press, 1999.

[21] A. J. Paulraj, D. Gore, R. U. Nabar, and H. Boelcskei, "An overview of MIMO communications-A key to Gigabit wireless," *Proc. IEEE*, vol. 92, no. 2, pp. 198–218, Feb. 2004.

[22] Y. Rong, S. A. Vorobyov, and A. B. Gershman, "Robust linear receivers for multi-access space-time block coded MIMO systems: A probabilistically constrained approach," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1560–1570, Aug. 2006.

[23] Y. Rong, S. A. Vorobyov, and A. B. Gershman, "A robust linear receiver for multi-access space-time block-coded MIMO systems based on probability-constrained optimization," *Proc. IEEE Vehic. Technol. Conf.*, vol. 1, pp. 118–122, May 2004.

[24] Y. Rong, S. Shahbazpanahi, and A. B. Gershman, "Robust linear receivers for space-time block coded multi-access MIMO systems with imperfect channel state information," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pt. 2, pp. 3081–3090, Aug. 2005.

[25] J. B. Schodorf and D. B. Williams, "A constrained optimization approach to multiuser detection," *IEEE Trans. Signal Process.*, vol. 45, no. 1, pp. 258–262, Jan. 1997.

[26] S. Shahbazpanahi, A. B. Gershman, Z. Luo, and K. M. Wong, "Robust adaptive beamforming for general-rank signal models," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2257–2269, Nov. 2004.

[27] J. R. Schott, *Matrix Analysis for Statistics*, 2nd ed. New York: Wiley, 2005.

[28] G. W. Stewart, "Error and perturbation bounds for subspaces associated with certain eigenvalue problems," *SIAM Rev.*, no. 15–33, pp. 1413–1416, 1973.

[29] G. L. Stuber, J. R. Barry, S. W. Mclaughlin, Y. Li, M. A. Ingram, and T. G. Pratt, "Broadband MIMO-OFDM wireless communications," *Proc. IEEE*, vol. 92, no. 2, pp. 271–294, Feb. 2004.

[30] Z. Tian, K. L. Bell, and H. L. Van Trees, "Robust constrained linear receivers for CDMA wireless systems," *IEEE Trans. Signal Process.*, vol. 49, no. 7, pp. 1510–1522, Jul. 2001.

[31] S. Trautmann and N. J. Fliege, "A new equalizer for multitone systems without guard time," *IEEE Comm. Lett.*, vol. 6, no. 1, pp. 34–36, Jan. 2002.
[32] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.*, vol. 5, pp. 4–24, Apr. 1988.
[33] S. A. Vorobyov, A. Gershman, and Z. Luo, "Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 313–324, Feb. 2003.
[34] H. L. Van Trees, *Optimum Array Processing*. New York: Wiley , 2002.
[35] Z. Wang and G. B. Giannakis, "Wireless multicarrier communications: Where Fourier meets Shannon," *IEEE Signal Process. Mag.*, vol. 17, no. 3, pp. 29–48, May 2000.
[36] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over rich scattering wireless channels," in *Proc. IEEE ISSSE*, Sep. 1998, pp. 295–300.
[37] Z. Xu, "Perturbation analysis for subspace decomposition with applications in subspace-based algorithms," *IEEE Trans. Signal Process.*, vol. 50, no. 11, pp. 2820–2830, Nov. 2002.
[38] K. Zarifi, S. Shahbazpanahi, A. B. Gershman, and Z. Luo, "Robust blind multiuser detection based on the worse-case performance optimization o the MMSE receiver," *IEEE Trans. Signal Process.*, vol. 53, no. 1, pp. 295–305, Jan. 2005.

Chih-Yuan Lin was born in Yilan, Taiwan, R.O.C., in 1979. He received the B.S. degree from National Chung Hsing University, Taichung, Taiwan, in 2000. He is currently working toward the Ph.D. degree in the Department of Communication Engineering at National Chiao Tung University, Hsinchu, Taiwan.

His current research interests include space-time signal processing for wireless communications and statistical signal processing.

Jwo-Yuh Wu (M'04) received the B.S. degree in 1996, the M.S. degree in 1998, and the Ph.D. degree in 2002 from National Chiao Tung University, Hsinchu, Taiwan, R.O.C, all in electrical and control engineering.

He is currently a Postdoctoral Research Fellow in the Department of Communication Engineering, National Chiao Tung University. His research interests are in signal processing and information theory, with current emphasis on communications and networking.

Ta-Sung Lee (M'94–SM'05) was born in Taipei, Taiwan, R.O.C., in 1960. He received the B.S. degree from National Taiwan University, Taipei, Taiwan, R.O.C., in 1983, the M.S. degree from the University of Wisconsin, Madison, in 1987, and the Ph.D. degree from Purdue University, West Lafayette, IN, in 1989, all in electrical engineering.

In 1990, he joined the Faculty of National Chiao Tung University (NCTU), Hsinchu, Taiwan, where he holds a position as Professor and Chairman of Department of Communication Engineering. His other positions include Technical Advisor at Information and Communications Research Laboratories (CCL), Industrial Technology Research Institute (ITRI), Taiwan, Managing Director of MINDS Research Center, College of Electrical and Computer Engineering, NCTU, and Managing Director of Communications and Computer Training Program, NCTU. He is active in research and development in advanced techniques for wireless communications, such as smart antenna and MIMO technologies, cross-layer design, and SDR prototyping of advanced communication systems. He has co-led several National Research Programs, such as the Program for Promoting Academic Excellence of Universities—Phases I & II and the 4G Mobile Communications Research Program Sponsored by Taiwan Government.

Dr. Lee has won several awards for his research, engineering, and teaching contributions; these include two times the National Science Council (NSC) Superior Research Award, the 1999 Young Electrical Engineer Award of the Chinese Institute of Electrical Engineers, and the 2001 NCTU Teaching Award.