Adaptive Decision Feedback Space–Time Equalization With Generalized Sidelobe Cancellation

Yinman Lee, *Member, IEEE*, and Wen-Rong Wu, *Member, IEEE*

***Abstract*—In wireless communications, cochannel interference (CCI) and intersymbol interference (ISI) are two main factors that limit system performance. Conventionally, a beamformer is used to reduce CCI, whereas an equalizer is used to compensate for ISI. These two devices can be combined into one as space–time equalizer (STE). A training sequence is usually required to train the STE prior to its use. In some applications, however, spatial information corresponding to a desired user is available, but the training sequence is not. In this paper, we propose an adaptive decision feedback STE to cope with this problem. Our scheme consists of an adaptive decision feedback generalized sidelobe canceller (DFGSC), a blind decision feedback equalizer (DFE), and a channel estimator. Due to the feedback operation, the proposed DFGSC is not only superior to the conventional generalized sidelobe canceller but also robust to multipath channel propagation and spatial signature error. Theoretical results are derived for optimum solutions, convergence behavior, and robustness properties. With the special channel-aided architecture, the proposed blind DFE can reduce the error propagation effect and be more stable than the conventional blind DFE. Simulation results show that the proposed STE is effective in mitigating both CCI and ISI, even in severe channel environments.**

***Index Terms*—Blind equalization, channel estimation, decision feedback equalizer (DFE), generalized sidelobe canceller (GSC), space–time equalizer (STE).**

I. Introduction

IN WIRELESS communications, the cochannel interference (CCI) due to multiple access and the intersymbol interference (ISI) due to multipath channels often cause severe signal distortion and limit system performance [1], [2]. In recent years, there has been a growing interest in applying adaptive antenna arrays and space–time signal processing techniques to solve these problems [3]–[5]. The common approach is to use a beamformer for CCI reduction and an equalizer for ISI compensation. These two devices can further be combined into one as space–time equalizer (STE) [6], [7]. The application of the STE is beneficial to communication quality, subsequently improving the detection performance, even in severe channel environments.

The optimum STE is known to be a maximum likelihood sequence estimator (MLSE) operated in the space–time domain [7]. However, the MLSE is notorious for high computational complexity. A suboptimum approach with a hybrid of a linear filter and a Viterbi equalizer was proposed in [8]. Even so, the implementation complexity is still high, and it limits the MLSE-like structure in real-world applications. The STE performing both beamforming and equalization, as aforementioned, is what researchers consider most. The general form of the STE consists of an antenna array and a temporal filter bank [9], [10]. Either a linear equalizer or a decision feedback equalizer (DFE) can be applied to the structure. Although the performance of this structure is satisfactory, its computational complexity is quite high. The other problem is that when an adaptive algorithm is applied, the convergence is slow, particularly operating under a large number of antenna elements and a severe fading channel. To ease these problems, another form that is a hybrid (or cascade) of a spatial filter and a temporal filter was proposed for the STE [11], [12]. It requires lower computational complexity, and the convergence is faster. However, the space–time information of the received signal will not be fully exploited, and there will be some performance loss.

All the aforementioned STEs are training based. In other words, we have to transmit training sequences before actually using them. It is well known that the transmission of training sequences will reduce the bandwidth utilization efficiency and may not always be possible. In some applications, however, spatial information corresponding to a desired user is available. For instance, a receiver may perform direction of arrival (DOA) estimation before signal detection. Moreover, a space division multiple access (SDMA) system employed to increase system capacity [13]–[16] transmits or receives a signal only from a certain direction. A practical example is the broadband wireless access system, particularly for fixed wireless applications. In these cases, it is possible to utilize the *a priori* spatial information and avoid the requirement of training sequences. A straightforward idea is to use a generalized sidelobe canceller (GSC) [17] for CCI suppression and a blind equalizer for ISI compensation. Unfortunately, the result of this direct cascade is often far from satisfactory. The reasons are stated in the subsequent sections.

Manuscript received February 5, 2007; revised July 26, 2007, October 8, 2007, and November 12, 2007. This work was supported by the National Science Council, Taiwan, R.O.C., under Grant NSC 96-2219-E-009-018. The review of this paper was coordinated by Prof. J. Choi.

Y. Lee is with the Graduate Institute of Communication Engineering, National Chi Nan University, Nantou 545, Taiwan, R.O.C. (e-mail: ymlee@ncnu.edu.tw).

W.-R. Wu is with the Department of Communication Engineering, National Chiao Tung University, Hsinchu 300, Taiwan, R.O.C. (e-mail: wrwu@faculty.nctu.edu.tw).

Digital Object Identifier 10.1109/TVT.2007.914472

For computational complexity consideration, the GSC is often implemented with an adaptive structure. The least-mean-square (LMS) algorithm is a well-known and widely used adaptive algorithm [18]. However, the LMS-based adaptive GSC usually converges slowly. The error signal for the LMS adaptation in this case consists of the desired signal component, and so, it is large, even in a noiseless environment. This large error signal significantly magnifies the mean-squared error (MSE). To reduce the MSE, the step size (a parameter controlling the LMS convergence) must be small. This essentially makes the LMS algorithm converge slowly. In addition, the adaptive GSC is sensitive to constraint mismatch, which is caused by incorrect spatial information. In this case, the signal cancellation phenomenon will occur, and the performance of the adaptive GSC can seriously be degraded. In typical applications, constraint mismatch can easily arise due to multipath channels and spatial signature errors.

It is well known that the DFE can have much better performance than the linear equalizer in severe fading channels. This statement is also true for blind equalization. However, the blind DFE is difficult to derive and rarely reported in the literature. The major contribution in this field is from Labat *et al.*, who proposed an interesting blind DFE in [19]. They used an infinite impulse response (IIR) whitening filter cascaded with a blind finite impulse response (FIR) linear equalizer in the startup period. After convergence, it switches the cascading order of the IIR and FIR filters, yielding a decision feedback structure. At the same time, a decision-directed minimum MSE (MMSE) training is initiated. For easy reference, we call this blind DFE the LBDFE hereafter. One inherent problem associated with the DFE is its error propagation effect, and this will have even more impact in its adaptive implementation. Since the LBDFE uses decision-directed training, it is sensitive to error propagation.

In this paper, we propose an adaptive STE for systems with *a priori* spatial information. The proposed structure comprises a decision feedback GSC (DFGSC), a blind DFE, and a channel estimator. The DFGSC structure can eliminate the desired signal component from the error signal in the LMS adaptation. As a consequence, it not only improves CCI suppression but also allows the simple point distortionless constraint to be robust to multipath channel environments and spatial signature errors. The proposed blind DFE adapts a channel-aided structure, yielding better ISI compensation and higher resistance to error propagation. We will demonstrate that the proposed blind DFE performs better than the LBDFE. This paper is an extension of [20], in which only additive white Gaussian noise (AWGN) channel was considered.

This paper is organized as follows. In Section II, the space–time signal model for the STE is developed. The effect of both CCI and ISI on the desired signal is also explained. In Section III, we describe a straightforward approach of the conventional GSC and LBDFE in a hybrid manner. In Section IV, we propose the new adaptive STE and describe the corresponding operation mechanisms in detail. Section V shows that the proposed STE is robust to general space–time multipath channels and spatial signature errors. Finally, simulation results and conclusions are given in Sections VI and VII, respectively.

II. Space–Time Signal Model

Consider a uniform linear array (ULA) of N antenna elements at the receiver. Let the $N \times 1$ received equivalent complex baseband signal vector in continuous time be denoted by

$$\mathbf{x}(t) = [x_0(t) \quad x_1(t) \quad \cdots \quad x_{N-1}(t)]^T \tag{1}$$

where $x_n(t)$ ($0 \le n \le N-1$, for all n hereafter) is the signal received from the nth antenna element at time t, and the superscript $(\cdot)^T$ denotes the transpose operation. Define an $N \times 1$ vector $\mathbf{a}(\theta)$ as the spatial signature for the signal from the DOA θ. It is written as $\mathbf{a}(\theta) = [1 \; e^{i\zeta_\theta} \; e^{i2\zeta_\theta} \cdots e^{i(N-1)\zeta_\theta}]^T$, where $i = \sqrt{-1}$, and $\zeta_\theta = (2\pi d/L_\lambda)\sin\theta$, in which d is the element spacing, and L_λ is the signal wavelength. Assume that there are M sources (including both desired and interfering sources) coming from M different and distinguishable directions. The transmitted signal waveform of the mth source $s_m(t)$ ($0 \le m \le M-1$, for all m hereafter) can be written as

$$s_m(t) = \sum_{k=-\infty}^{+\infty} p(t - kT) b_m(k) \tag{2}$$

where $b_m(k)$ is the kth complex information symbol of the mth source, $p(t)$ is the pulse shape of the transmitted symbol, and T is the symbol duration. Let L_m be the number of propagation paths for the mth source, θ_m be the DOA for the mth source, τ_{ml} be the delay time for the lth path ($0 \le l \le L_m - 1$, for all l hereafter) of the mth source, and $\bar{h}_{ml}$ be the complex channel coefficient for the lth path of the mth source, respectively. Assume that the channel parameters for different sources are independent and remain constant over the observation period. The received signal vector in (1) can then be expressed as

$$\mathbf{x}(t) = \sum_{m=0}^{M-1} \sum_{l=0}^{L_m-1} \bar{h}_{ml} s_m(t - \tau_{ml}) \mathbf{a}(\theta_m) + \mathbf{n}(t) \tag{3}$$

where $\mathbf{n}(t)$ is an $N \times 1$ complex AWGN vector. Each component in the AWGN vector is spatially white with a variance of σ_n^2. Substituting (2) into (3), we have

$$\mathbf{x}(t) = \sum_{m=0}^{M-1} \sum_{k=-\infty}^{+\infty} \mathbf{h}_m(t - kT) b_m(k) + \mathbf{n}(t) \tag{4}$$

where $\mathbf{h}_m(t)$ is an $N \times 1$ vector summarizing the total transmission effect of the mth source on the information, and it is given by

$$\mathbf{h}_m(t) = \sum_{l=0}^{L_m-1} \bar{h}_{ml} p(t - \tau_{ml}) \mathbf{a}(\theta_m). \tag{5}$$

Without loss of generality, we assume that the first M_0 sources ($M_0 \le M$) are related to the desired user, i.e., $b_0(k) = b_1(k) = \cdots = b_{M_0-1}(k)$. The DOA θ_0 is the DOA of the desired signal with the strongest signal strength, and θ_1 to θ_{M_0-1} correspond to the DOAs of the desired signal's spatial ISI. The received equivalent complex baseband signal is sampled at the symbol rate, i.e., $t = kT$. The sampling clock is synchronized with the transmission clock. After sampling, the channel effect for all

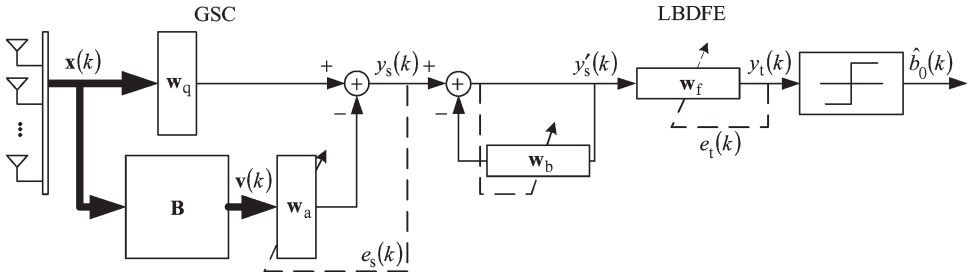


Fig. 1. Hybrid of GSC and LBDFE (with LBDFE in startup period).

sources is of a finite duration within $[0,(D_m-1)T]$, where D_m is the channel order of the mth source. Here, we define $\mathbf{d}(k)$ as an $N\times 1$ vector representing the components from the main source of the desired signal $(m=0)$, and $\mathbf{i}(k)$ as an $N\times 1$ vector summing up the components from the spatial ISI sources $(m=1,2,\ldots,M_0-1)$, and they are, respectively, given by

$$\mathbf{d}(k)=\sum_{d=0}^{D_0-1}\mathbf{h}_0(d)b_0(k-d) \tag{6}$$

$$\mathbf{i}(k)=\sum_{m=1}^{M_0-1}\sum_{d=0}^{D_m-1}\mathbf{h}_m(d)b_0(k-d). \tag{7}$$

In addition, define an $N\times 1$ vector $\mathbf{z}(k)$ representing the uncorrelated CCI-plus-noise components as

$$\mathbf{z}(k)=\sum_{m=M_0}^{M-1}\sum_{d=0}^{D_m-1}\mathbf{h}_m(d)b_m(k-d)+\mathbf{n}(k). \tag{8}$$

Then, the expression in (4) can be rewritten as

$$\mathbf{x}(k)=\mathbf{d}(k)+\mathbf{i}(k)+\mathbf{z}(k). \tag{9}$$

We use this signal model to describe various kinds of CCI and ISI in the wireless communication environment. The main task of the STE is to suppress interference and recover the transmitted information. To simplify the notations, we write $b(k)$ instead of $b_0(k)$ for the desired user's information symbols in the following derivation.

III. Hybrid of GSC and LBDFE

In some wireless communication systems, such as SDMA applications, the main DOA (and hence the spatial signature) of the desired signal is known *a priori* or can be estimated. A straightforward approach, as mentioned before for CCI and ISI mitigation, is the hybrid of a conventional GSC and an LBDFE. The purpose of the conventional GSC is to suppress CCI, whereas that of the LBDFE is to compensate for ISI. With the spatial signature, no extra training sequence is required for either processing. The operation and weakness of this approach will subsequently be elaborated upon.

A. GSC

The conventional GSC for CCI suppression is optimized with the linearly constrained minimum variance (LCMV) criterion [21]. The LCMV beamformer determines the N-tap weight vector $\mathbf{w}$ through

$$\min_{\mathbf{w}}\quad \mathbf{w}^H\mathbf{R}_{\mathbf{x}}\mathbf{w}\quad\text{subject to}\quad \mathbf{C}^H\mathbf{w}=\mathbf{f} \tag{10}$$

where $\mathbf{R}_{\mathbf{x}}=\mathrm{E}\{\mathbf{x}(k)\mathbf{x}^H(k)\}$ is the input correlation matrix, $\mathbf{C}$ is an $N\times U$ constraint matrix, and $\mathbf{f}$ is a $U\times 1$ response vector, with U being the number of constraints. The superscript $(\cdot)^H$ denotes Hermitian transposition. With the structure of GSC, as shown in the left part of Fig. 1, the constrained optimization problem can be transformed into an unconstrained optimization problem [17]. This structure can effectively reduce the computational cost, particularly when implemented with adaptive algorithms. As illustrated in the figure, the upper path includes an N-tap quiescent signal matched filter $\mathbf{w}_q$. The lower path includes an $N\times(N-U)$ blocking matrix $\mathbf{B}$ and an $(N-U)$-tap interference canceling filter $\mathbf{w}_a$. Then, we have an equivalent spatial filter as $\mathbf{w}=\mathbf{w}_q-\mathbf{B}\mathbf{w}_a$. Ideally, the span of $\mathbf{B}$ is in the null space of $\mathbf{C}^H$. Using the constraint in (10), $\mathbf{w}_q$ can readily be found to be $\mathbf{w}_q=\mathbf{C}(\mathbf{C}^H\mathbf{C})^{-1}\mathbf{f}$, and $\mathbf{w}_a$ is optimized according to the output power of the GSC as

$$J=\mathrm{E}\left\{|y_s(k)|^2\right\}=\mathrm{E}\left\{\left|(\mathbf{w}_q-\mathbf{B}\mathbf{w}_a)^H\mathbf{x}(k)\right|^2\right\}. \tag{11}$$

The constrained optimization problem in (10) can then be rewritten as the following unconstrained optimization problem:

$$\min_{\mathbf{w}_a}\quad J=\min_{\mathbf{w}_a}\quad(\mathbf{w}_q-\mathbf{B}\mathbf{w}_a)^H\mathbf{R}_{\mathbf{x}}(\mathbf{w}_q-\mathbf{B}\mathbf{w}_a). \tag{12}$$

The optimum $\mathbf{w}_a$ is classically solved to be [17]

$$\mathbf{w}_{a,\mathrm{opt}}=(\mathbf{B}^H\mathbf{R}_{\mathbf{x}}\mathbf{B})^{-1}\mathbf{B}^H\mathbf{R}_{\mathbf{x}}\mathbf{w}_q. \tag{13}$$

With the optimum weight vector, the minimum value of J in (11), which is denoted as $J_{\min}$ (which equals the minimum output power, which is denoted as $P_{o,\min}$ for the conventional GSC), can be calculated as

$$J_{\min}=P_{o,\min} \tag{14}$$

$$=\mathbf{w}_{\mathrm{opt}}^H\mathbf{R}_{\mathbf{x}}\mathbf{w}_{\mathrm{opt}} \tag{15}$$

$$=\mathbf{w}_q^H\mathbf{R}_{\mathbf{x}}\mathbf{w}_{\mathrm{opt}} \tag{16}$$

where we let $\mathbf{w}_{\mathrm{opt}}=\mathbf{w}_q-\mathbf{B}\mathbf{w}_{a,\mathrm{opt}}$. Since $\mathbf{C}$ and $\mathbf{f}$ are design parameters, $\mathbf{w}_q$ and $\mathbf{B}$ can be calculated offline. The calculation of $\mathbf{w}_{a,\mathrm{opt}}$ shown in (13), however, is much more involved. An alternative to find it is to use the adaptive training method. The LMS algorithm, being one of the stochastic gradient methods,

is chosen here for its simple yet effective nature. Taking the stochastic gradient of J with respect to $\mathbf{w}_a^*$, where $(\cdot)^*$ denotes the conjugate operation, we can obtain the LMS update equation for $\mathbf{w}_a$ as

$$\mathbf{w}_a(k+1) = \mathbf{w}_a(k) + \mu_a \mathbf{v}(k) e_s^*(k) \tag{17}$$

where $\mathbf{w}_a(k)$ is the estimate of $\mathbf{w}_{a,\mathrm{opt}}$ at the kth iteration, $\mathbf{v}(k) = \mathbf{B}^H \mathbf{x}(k)$ is the filter input vector (the output vector from the blocking matrix), μ_a is the step size controlling the convergence rate, and $e_s(k)$ is an error signal between the desired and actual output. For GSC applications, we have $e_s(k) = y_s(k)$. When $\mathbf{w}_a$ is optimized, the error signal $e_s(k)$ will chiefly include the component from the desired signal, i.e., the desired user's transmitted information symbols $b(k)$. This indicates that the stochastic gradient in (17), i.e., $\mathbf{v}(k) y_s^*(k)$, will be large, even for optimum weights. When the LMS algorithm is applied to estimate $\mathbf{w}_{a,\mathrm{opt}}$, the performance will be affected due to the large error signal used. The other problem with the conventional GSC is its sensitivity to constraint mismatch. Whenever the setting of the constraint matrix $\mathbf{C}$ in (10) [or the blocking matrix $\mathbf{B}$ in (12)] is not fit for the actual spatial signature of the desired signal, constraint mismatch occurs. Constraint mismatch can easily arise due to multipath channels and spatial signature errors. This seriously degrades the performance of the adaptive GSC. All these problems will be analyzed in depth later.

B. LBDFE

The output of the GSC is fed into the LBDFE for equalization. The equalizer adaptation process is divided into two periods. In the startup period, the received signal is prewhitened by an IIR filter and then equalized by a blind FIR linear equalizer. The structure of the LBDFE in the startup period is shown in the right part of Fig. 1. In [19], the constant modulus algorithm (CMA) was used as the blind algorithm for equalization. Recently, a sophisticated blind equalization algorithm called the multimodulus algorithm (MMA) was proposed [22]. Analysis shows that the MMA can provide much more stable performance, particularly with high-order constellation modulation [23]. The cost function can yield an equalized constellation rotated with a multiple of 90°, eliminating the need for additional constellation phase recovery as needed in the CMA. The remaining phase ambiguity problem is classically solved by differential encoding. For these reasons, we use the MMA as our blind equalization algorithm (instead of the CMA) throughout this paper. For the LBDFE, we denote the length of $\mathbf{w}_f$ and $\mathbf{w}_b$ as α and β, respectively. In the startup period, let $y_s(k)$ be the output of the GSC, i.e., $y_s(k) = (\mathbf{w}_q - \mathbf{B}\mathbf{w}_a)^H \mathbf{x}(k)$, as given in (11), and let $y_s'(k)$ be the difference between $y_s(k)$ and the prewhitening filter output. In addition, let $\mathbf{y}'(k)$ be the input vector of $\mathbf{w}_b$ as $\mathbf{y}'(k) = [y_s'(k-1) \;\; y_s'(k-2) \;\; \cdots \;\; y_s'(k-\beta)]^T$. So, we have $y_s'(k) = y_s(k) - \mathbf{w}_b^H \mathbf{y}'(k)$. The filter $\mathbf{w}_b$ is optimized according to the criterion $\min_{\mathbf{w}_b} \; \mathrm{E}\{|y_s'(k)|^2\}$. As shown in Fig. 1, $y_s'(k)$ serves as the input of the blind FIR filter $\mathbf{w}_f$ as well. We define another vector as $\mathbf{y}''(k) = [y_s'(k) \;\; y_s'(k-1) \;\; \cdots \;\; y_s'(k-\alpha+1)]^T$, and, as a consequence, the output of the LBDFE in the startup period is $y_t(k) = \mathbf{w}_f^H \mathbf{y}''(k)$. For the MMA, the error signal is defined as

$$e_t(k) = \left(y_{t,r}^2(k) - R^2\right)^2 + \left(y_{t,i}^2(k) - R^2\right)^2 \tag{18}$$

with

$$R^2 = \frac{\mathrm{E}\left\{b_r^4(k)\right\}}{\mathrm{E}\left\{b_r^2(k)\right\}} = \frac{\mathrm{E}\left\{b_i^4(k)\right\}}{\mathrm{E}\left\{b_i^2(k)\right\}} \tag{19}$$

in which the subscripts r and i in $y_t(k)$ and $b(k)$ denote the real and imaginary parts of $y_t(k)$ and $b(k)$, respectively. The cost function for the optimization of $\mathbf{w}_f$ is then written as

$$\begin{aligned} J_{\mathrm{MMA}} &= \mathrm{E}\left\{e_t(k)\right\} \\ &= \mathrm{E}\left\{\left(y_{t,r}^2(k) - R^2\right)^2 + \left(y_{t,i}^2(k) - R^2\right)^2\right\}. \end{aligned} \tag{20}$$

Let μ_f be the step size controlling the convergence of $\mathbf{w}_f$. From [22], the stochastic gradient algorithm for the MMA of $\mathbf{w}_f$ can be obtained as

$$\mathbf{w}_f(k+1) = \mathbf{w}_f(k) + \mu_f \phi\left(y_t(k)\right) \mathbf{y}''(k) \tag{21}$$

with

$$\phi\left(y_t(k)\right) = y_{t,r}^3(k) + i y_{t,i}^3(k) - R^2 y_t(k). \tag{22}$$

The startup period is expected to sufficiently open the eye pattern such that the error rate is low enough to initiate the second (tracking) period. In the tracking period, the cascading order of the IIR prewhitening filter and the FIR linear equalizer is swapped, and this turns the whole system into a conventional DFE structure. A decision-directed MMSE tracking operation, similar to that of the conventional DFE, is then activated. This approach may initially avoid the possible error propagation phenomenon and give a smooth transition strategy between blind and decision-directed equalization. Nevertheless, the stability of the adaptive IIR filter makes the performance sensitive to the parameters chosen. In addition, error propagation may occur during the decision-directed mode. The behavior of the LBDFE becomes not easy to control. We will empirically show in Section VI that the performance of the hybrid of the conventional GSC and LBDFE is often far from satisfactory in severe channel environments.

IV. Proposed Hybrid STE

In this section, we propose a new adaptive STE that is a hybrid of DFGSC and a channel-aided blind DFE (CBDFE). Fig. 2 shows the block diagram of the proposed STE. Note that in Fig. 2, a channel estimator is included. Let the coefficients of the channel estimator be denoted as $\mathbf{w}_h$. It is used to model the equivalent temporal channel for the desired signal, and its operation will be explained soon. In [11], a channel estimator was also used in a training-based STE such that the corresponding beamformer can achieve better performance. In the STE scheme proposed here, the role of the channel estimator is much more involved. For spatial processing, with the help of the channel estimator, we can formulate the adaptive DFGSC,

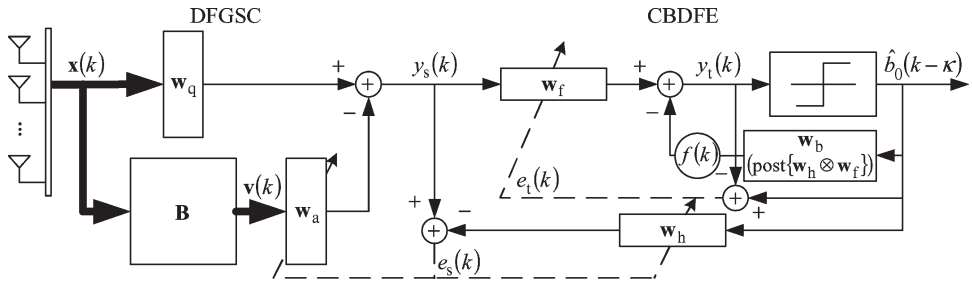


Fig. 2. Proposed hybrid STE.

achieving better CCI suppression performance. In addition, the DFGSC can have extra robustness against constraint mismatch. For temporal processing, with the help of the channel estimator, we can formulate the CBDFE such that it is more effective in ISI compensation. This CBDFE is different from those channel-estimation-based DFEs proposed in [24] and [25], where the feedforward and feedback filters are calculated based on the estimated channel response. In the CBDFE, however, the adaptive structure is preserved. The operation of the DFGSC and CBDFE is separately presented as follows.

A. Proposed DFGSC

Let the channel estimator $\mathbf{w}_h$ have a dimension of $\gamma \times 1$. Here, the value of γ is chosen to be equal to or larger than the maximum value of the channel order D_m $(0 \le m \le M_0 - 1)$. We first define $\hat{\mathbf{b}}(k)$ as the input vector of $\mathbf{w}_h$, i.e., $\hat{\mathbf{b}}(k) = [\hat{b}(k) \quad \hat{b}(k-1) \quad \cdots \quad \hat{b}(k-\gamma+1)]^T$. To optimize the interference canceling filter $\mathbf{w}_a$ and the channel estimator $\mathbf{w}_h$, we propose a new cost function as

$$\begin{aligned} J &= \mathrm{E}\left\{|e_s(k)|^2\right\} \\ &= \mathrm{E}\left\{\left|y_s(k) - \mathbf{w}_h^H \hat{\mathbf{b}}(k)\right|^2\right\} \\ &= \mathrm{E}\left\{\left|\mathbf{w}_q^H \mathbf{x}(k) - \begin{bmatrix}\mathbf{w}_a^H & \mathbf{w}_h^H\end{bmatrix}\begin{bmatrix}\mathbf{B}^H\mathbf{x}(k) \\ \hat{\mathbf{b}}(k)\end{bmatrix}\right|^2\right\}. \end{aligned} \tag{23}$$

To understand the operation mechanisms, we first study a simplified scenario in which only the temporal ISI is present. In other words, the desired signal and its ISI only come from the main DOA, i.e., $M_0 = 1$ (and so $\mathbf{i}(k)$ in (7) is a zero vector). Consider that the spatial signature for the desired signal is exactly known, and a distortionless constraint is set toward it. Thus, $U = 1$ is used in the derivation. In this case, the desired signal will be completely blocked in the lower path of the DFGSC. The general case where spatial ISI exists and the distortionless constraint may be set improperly will be discussed in the next section. Assuming that the decision is correct, i.e., $\hat{b}(k) = b(k)$, we can rewrite the minimization of the cost function in (23) as

$$\begin{aligned} \min_{\mathbf{w}_a, \mathbf{w}_h} J = \min_{\mathbf{w_c}} & \mathbf{w}_q^H \mathbf{R}_\mathbf{x} \mathbf{w}_q - \mathbf{w}_q^H [\mathbf{R}_\mathbf{x}\mathbf{B} \quad \mathbf{R}_\mathbf{p}]\mathbf{w_c} \\ & - \mathbf{w_c}^H \begin{bmatrix}\mathbf{B}^H\mathbf{R}_\mathbf{x} \\ \mathbf{R}_\mathbf{p}^H\end{bmatrix}\mathbf{w}_q + \mathbf{w_c}^H \mathbf{R_c} \mathbf{w_c} \end{aligned} \tag{24}$$

where we let

$$\mathbf{w_c} = \begin{bmatrix}\mathbf{w}_a \\ \mathbf{w}_h\end{bmatrix} \tag{25}$$

$$\begin{aligned} \mathbf{R_c} &= \mathrm{E}\left\{\begin{bmatrix}\mathbf{B}^H\mathbf{x}(k) \\ \hat{\mathbf{b}}(k)\end{bmatrix}\begin{bmatrix}\mathbf{x}^H(k)\mathbf{B} & \hat{\mathbf{b}}^H(k)\end{bmatrix}\right\} \\ &= \begin{bmatrix}\mathbf{B}^H\mathbf{R}_\mathbf{x}\mathbf{B} & \mathbf{0} \\ \mathbf{0}^H & \sigma_b^2\mathbf{I}_\gamma\end{bmatrix} \end{aligned} \tag{26}$$

$$\begin{aligned} \mathbf{R_p} &= \mathrm{E}\left\{\mathbf{x}(k)\hat{\mathbf{b}}^H(k)\right\} \\ &= \sigma_b^2 [\mathbf{h}_0(0) \quad \mathbf{h}_0(1) \quad \cdots \quad \mathbf{h}_0(\gamma-1)] \end{aligned} \tag{27}$$

in which $\mathbf{0}$ denotes a zero matrix with dimension $(N-1) \times \gamma$, $\mathbf{I}_\gamma$ denotes an identity matrix with dimension $\gamma \times \gamma$, and σ_b^2 denotes the power of transmitted symbols. Since there is no correlation between $\mathbf{B}^H\mathbf{x}(k)$ and $\hat{\mathbf{b}}(k)$ in this case, the off-diagonal block of $\mathbf{R_c}$ in (26) is zero. Again, $\mathbf{w}_q$ is for signal matching, and it can be set the same as for the conventional GSC. Taking the derivative of this cost function with respect to $\mathbf{w_c}^*$ and setting the result to zero, we can obtain the optimum $\mathbf{w_c}$ as

$$\frac{\partial J}{\partial \mathbf{w_c}^*} = -2\begin{bmatrix}\mathbf{B}^H\mathbf{R}_\mathbf{x} \\ \mathbf{R}_\mathbf{p}^H\end{bmatrix}\mathbf{w}_q + 2\mathbf{R_c}\mathbf{w_c} = 0. \tag{28}$$

Thus

$$\mathbf{w}_{\mathbf{c},\mathrm{opt}} = \mathbf{R_c}^{-1}\begin{bmatrix}\mathbf{B}^H\mathbf{R}_\mathbf{x} \\ \mathbf{R}_\mathbf{p}^H\end{bmatrix}\mathbf{w}_q. \tag{29}$$

With the special structure of $\mathbf{R_c}$, we can decompose $\mathbf{w}_{\mathbf{c},\mathrm{opt}}$ back into the two weights

$$\mathbf{w}_{a,\mathrm{opt}} = (\mathbf{B}^H\mathbf{R}_\mathbf{x}\mathbf{B})^{-1}\mathbf{B}^H\mathbf{R}_\mathbf{x}\mathbf{w}_q \tag{30}$$

$$\begin{aligned} \mathbf{w}_{h,\mathrm{opt}} &= \frac{\mathbf{R}_\mathbf{p}^H}{\sigma_b^2}\mathbf{w}_q \\ &= [\mathbf{h}_0(0) \quad \mathbf{h}_0(1) \cdots \mathbf{h}_0(\gamma-1)]^H \mathbf{w}_q. \end{aligned} \tag{31}$$

From (30), we observe that the expression of $\mathbf{w}_{a,\mathrm{opt}}$ is the same as that of the conventional GSC given in (13). Since $\mathbf{w}_q$ and $\mathbf{w}_{a,\mathrm{opt}}$ remain the same, the minimum output power $P_{o,\min}$ for the DFGSC is also the same. From (31), we observe that $\mathbf{w}_{h,\mathrm{opt}}$ corresponds to the equivalent channel effect from the source to

the DFGSC output. With $\mathbf{w}_{a,\mathrm{opt}}$ and $\mathbf{w}_{h,\mathrm{opt}}$ given earlier, the minimum J in (23) for the DFGSC becomes

$$\begin{aligned} J_{\min} &= \mathrm{E}\left\{\left|\mathbf{w}_{\mathrm{opt}}^H\mathbf{x}(k)-\mathbf{w}_{h,\mathrm{opt}}^H\hat{\mathbf{b}}(k)\right|^2\right\} \\ &= \mathbf{w}_q^H\mathbf{R}_{\mathbf{x}}\mathbf{w}_{\mathrm{opt}}-\sigma_b^2\left|\mathbf{w}_q^H[\mathbf{h}_0(0)\quad\mathbf{h}_0(1)\quad\cdots\quad\mathbf{h}_0(\gamma-1)]\right|^2 \\ &= P_{o,\min}-\sigma_b^2\left|\mathbf{w}_q^H[\mathbf{h}_0(0)\quad\mathbf{h}_0(1)\quad\cdots\quad\mathbf{h}_0(\gamma-1)]\right|^2. \end{aligned} \tag{32}$$

Note here that $J_{\min}$ is no longer equal to $P_{o,\min}$. We let $\mathbf{R}_{\mathbf{d}} = \mathrm{E}\{\mathbf{d}(k)\mathbf{d}^H(k)\}$ be the input correlation matrix of the desired signal excluding spatial ISI and $\mathbf{R}_{\mathbf{z}} = \mathrm{E}\{\mathbf{z}(k)\mathbf{z}^H(k)\}$ be the input correlation matrix of CCI-plus-noise. The first term in (32) is equal to the total power in the DFGSC output, and the second term in (32) is the power of the desired signal in the DFGSC output, which can be defined as

$$\begin{aligned} P_{\mathbf{d}} &\triangleq \mathbf{w}_{\mathrm{opt}}^H\mathbf{R}_{\mathbf{d}}\mathbf{w}_{\mathrm{opt}} \\ &= \sigma_b^2\left|\mathbf{w}_q^H\left[\mathbf{h}_0(0)\quad\mathbf{h}_0(1)\quad\cdots\quad\mathbf{h}_0(\gamma-1)\right]\right|^2. \end{aligned} \tag{33}$$

So, $J_{\min}$ in (32) can be calculated as

$$J_{\min} = P_{o,\min} - P_{\mathbf{d}} \tag{34}$$

$$= \mathbf{w}_q^H\mathbf{R}_{\mathbf{z}}\mathbf{w}_{\mathrm{opt}}. \tag{35}$$

Comparing (14) and (34), we can see that the desired signal is totally excluded with the help of $\mathbf{w}_{h,\mathrm{opt}}$. Thus, we conclude that the decision feedback only reduces the minimum J in (14), and the optimum performance is not enhanced. When an adaptive algorithm such as LMS is used to estimate $\mathbf{w}_{a,\mathrm{opt}}$, however, the performance of the GSC can greatly be improved by the feedback operation. The LMS update equations for the DFGSC can be written as

$$\mathbf{w}_a(k+1) = \mathbf{w}_a(k) + \mu_a\mathbf{v}(k)e_s^*(k) \tag{36}$$

$$\mathbf{w}_h(k+1) = \mathbf{w}_h(k) + \mu_h\hat{\mathbf{b}}(k)e_s^*(k) \tag{37}$$

where μ_a is the step size for $\mathbf{w}_a$, μ_h is the step size for $\mathbf{w}_h$, $\mathbf{v}(k)$ is the output from the blocking matrix, and $e_s(k) = y_s(k) - \mathbf{w}_h^H(k)\hat{\mathbf{b}}(k)$. Unlike the conventional GSC, the steady-state $e_s(k)$ will exclude the desired signal, and hence, it can be quite small. This is where the improvement of the adaptive DFGSC stems from.

To analyze the steady-state performance for both the conventional adaptive GSC and the proposed adaptive DFGSC, we denote the value of J [both (11) and (23)] in the steady state as $J(\infty)$. Suppose that the decision is correct, and $\mathbf{w}_h$ is fixed at optimum. Then

$$J(\infty) = J_{\min} + J_{\mathrm{ex}}(\infty) \tag{38}$$

where $J_{\mathrm{ex}}(\infty)$ is the excess MSE. Note that the excess MSE is yielded by the use of the LMS algorithm. In addition, define the weight error vector as

$$\boldsymbol{\epsilon}(k) = \mathbf{w}_a(k) - \mathbf{w}_{a,\mathrm{opt}}. \tag{39}$$

Using the direct averaging method [18], we can have

$$\boldsymbol{\epsilon}(k+1) = (\mathbf{I} - \mu_a\mathbf{B}^H\mathbf{R}_{\mathbf{x}}\mathbf{B})\boldsymbol{\epsilon}(k) + \mu_a\mathbf{B}^H\mathbf{x}(k)e_{\mathrm{opt}}^*(k) \tag{40}$$

where $e_{\mathrm{opt}}(k)$ denotes the error signal produced with the optimum weights. Define the correlation matrix of the weight error vector as

$$\mathbf{K}(k) = \mathrm{E}\left\{\boldsymbol{\epsilon}(k)\boldsymbol{\epsilon}^H(k)\right\}. \tag{41}$$

Invoking the independence assumption [18], we can obtain the recursive relation of $\mathbf{K}(k)$ as

$$\begin{aligned} \mathbf{K}(k+1) = (\mathbf{I} - \mu_a\mathbf{B}^H\mathbf{R}_{\mathbf{x}}\mathbf{B})\mathbf{K}(k)(\mathbf{I} - \mu_a\mathbf{B}^H\mathbf{R}_{\mathbf{x}}\mathbf{B}) \\ + \mu_a^2J_{\min}\mathbf{B}^H\mathbf{R}_{\mathbf{x}}\mathbf{B}. \end{aligned} \tag{42}$$

Under this premise, the excess MSE, which is denoted as $J_{\mathrm{ex}}(k)$, is written as

$$J_{\mathrm{ex}}(k) = \mathrm{tr}\left[\mathbf{B}^H\mathbf{R}_{\mathbf{x}}\mathbf{B}\mathbf{K}(k)\right] \tag{43}$$

where $\mathrm{tr}[\cdot]$ gives the trace of the matrix inside the brackets. As $k \to \infty$, $J_{\mathrm{ex}}(k)$ is given by

$$J_{\mathrm{ex}}(\infty) = J_{\min}\sum_{l=1}^{N-1}\frac{\mu_a\lambda_l(\mathbf{B}^H\mathbf{R}_{\mathbf{x}}\mathbf{B})}{2-\mu_a\lambda_l(\mathbf{B}^H\mathbf{R}_{\mathbf{x}}\mathbf{B})} \tag{44}$$

where $\lambda_l(\mathbf{B}^H\mathbf{R}_{\mathbf{x}}\mathbf{B})$ represents the lth eigenvalue of $\mathbf{B}^H\mathbf{R}_{\mathbf{x}}\mathbf{B}$. If μ_a is small, (44) can be approximated as

$$J_{\mathrm{ex}}(\infty) \approx \frac{\mu_aJ_{\min}}{2}\sum_{l=1}^{N-1}\lambda_l(\mathbf{B}^H\mathbf{R}_{\mathbf{x}}\mathbf{B}). \tag{45}$$

Thus, $J_{\mathrm{ex}}(\infty)$ is proportional to $J_{\min}$ and the step size μ_a. The transient output signal-to-interference-plus-noise ratio (SINR) for the conventional GSC and DFGSC can be written as

$$\mathrm{SINR}(k) = \frac{\mathrm{E}\left\{\left|\mathbf{w}^H(k)\mathbf{d}(k)\right|^2\right\}}{\mathrm{E}\left\{\left|\mathbf{w}^H(k)\mathbf{x}(k) - \mathbf{w}^H(k)\mathbf{d}(k)\right|^2\right\}}. \tag{46}$$

The optimum and steady-state output SINR (with the LMS algorithm) can then be calculated as

$$\begin{aligned} \mathrm{SINR}_{\mathrm{opt}} &= \frac{\mathrm{E}\left\{\left|\mathbf{w}_{\mathrm{opt}}^H\mathbf{d}(k)\right|^2\right\}}{\mathrm{E}\left\{\left|\mathbf{w}_{\mathrm{opt}}^H\mathbf{x}(k) - \mathbf{w}_{\mathrm{opt}}^H\mathbf{d}(k)\right|^2\right\}} \\ &= \frac{P_{\mathbf{d}}}{P_{o,\min} - P_{\mathbf{d}}} \end{aligned} \tag{47}$$

$$\begin{aligned} \mathrm{SINR}_{\mathrm{LMS}} &= \frac{\mathrm{E}\left\{\left|\mathbf{w}^H(\infty)\mathbf{d}(\infty)\right|^2\right\}}{\mathrm{E}\left\{\left|\mathbf{w}^H(\infty)\mathbf{x}(\infty) - \mathbf{w}^H(\infty)\mathbf{d}(\infty)\right|^2\right\}} \\ &= \frac{P_{\mathbf{d}}}{\mathbf{w}_q^H\mathbf{R}_{\mathbf{x}}\mathbf{w}_{\mathrm{opt}} + \mathrm{tr}\left[\mathbf{B}^H\mathbf{R}_{\mathbf{x}}\mathbf{B}\mathbf{K}(\infty)\right] - P_{\mathbf{d}}} \\ &= \frac{P_{\mathbf{d}}}{P_{o,\min} + J_{\mathrm{ex}}(\infty) - P_{\mathbf{d}}}. \end{aligned} \tag{48}$$

Note that the notations $\mathbf{w}(\infty)$, $\mathbf{d}(\infty)$, and $\mathbf{x}(\infty)$ are used to denote their final values and are based on the assumption of

reaching convergence under statistical expectation, just like that for $J_{\rm ex}(n) \to J_{\rm ex}(\infty)$ as $n \to \infty$. From (47), once again, we see that the optimum performance is not enhanced by the decision feedback operation since $P_{\mathbf{d}}$ and $P_{o,\min}$ are the same for the conventional GSC and the DFGSC. From (45), we can see that $J_{\rm ex}(\infty)$ is proportional to $J_{\min}$. The corresponding $J_{\min}$ values for both schemes are shown in (14) and (34), respectively. The resultant $J_{\min}$ for the adaptive DFGSC will be smaller than that for the conventional adaptive GSC. As a consequence, $J_{\rm ex}(\infty)$ for the adaptive DFGSC will be smaller than that for the conventional adaptive GSC. Moreover, from (48), we see that the smaller $J_{\rm ex}(\infty)$ is, the larger the steady-state output SINR we will have. We thus conclude that with the same step size, the steady-state output SINR of the adaptive DFGSC will be higher than that of the conventional adaptive GSC. In addition, note that the step size bound for the LMS algorithm is determined by the eigenvalue spread of the input correlation matrix. The input vectors for the conventional GSC and the DFGSC are the same, and so, the step size bounds for both schemes are the same.

B. Proposed CBDFE

The MMA is conventionally applied to the blind FIR linear equalizer only. Since the DFE is a nonlinear and IIR equalizer, direct application of the MMA may result in severe error propagation in the startup period. It may make the performance of the blind DFE even worse than that of its FIR linear companion. Here, we make use of the previously derived channel estimator $\mathbf{w}_h$ and propose a new structure, i.e., the CBDFE, to overcome this problem. In FIR linear equalization, the performance of the MMA can be demonstrated to approach that of the training-based MMSE method [19], [23]. Hence, we suppose that the weights solved by the MMA are close to those solved by the MMSE criterion. Our approach uses a basic principle of the DFE, i.e., the postcursor response of the channel convolved with the feedforward filter is cancelled by the feedback filter.[1] Let $\mathbf{w}_f$ and $\mathbf{w}_b$ be the feedforward and feedback filter weight vectors of a conventional DFE, with length α and β, respectively. Similar to the previous section, $y_s(k)$ denotes the output of the DFGSC. For notation simplicity, the input of $\mathbf{w}_f$ and $\mathbf{w}_b$ is again written as $\mathbf{y}(k) = [y_s(k) \quad y_s(k-1) \quad \cdots \quad y_s(k-\alpha+1)]^T$ and $\hat{\mathbf{b}}(k) = [\hat{b}(k-\kappa-1) \quad \hat{b}(k-\kappa-2) \quad \cdots \quad \hat{b}(k-\kappa-\beta)]^T$, where κ is the decision delay. Note that the convolution of the equivalent channel response and the feedforward filter results in a response of length $\alpha+\gamma-1$. Thus, for perfect postcursor cancellation, we must have $\beta \geq \alpha+\gamma-2-\kappa$. Without loss of generality, we let $\beta = \alpha+\gamma-2-\kappa$. We now prove the postcursor cancellation property mentioned earlier. Again, with correct decisions, the error signal is written as

$$\begin{aligned} e_t(k) &= b(k-\kappa) - y_t(k) \\ &= b(k-\kappa) - \left(\mathbf{w}_f^H \mathbf{y}(k) - \mathbf{w}_b^H \mathbf{b}(k)\right) \end{aligned} \tag{49}$$

[1]For DFE, precursor and postcursor responses are defined as the ISI from future and past symbols, respectively.

where $y_t(k)$ is the equalizer output. Straightforward manipulations give the equalizer output MSE as

$$\begin{aligned} \mathrm{E}\left\{|e_t(k)|^2\right\} = {} & \mathbf{w}_f^H \mathbf{R}_{\mathbf{yy}} \mathbf{w}_f - \mathbf{w}_f^H \mathbf{R}_{\mathbf{yb}} \mathbf{w}_b - \mathbf{w}_f^H \mathbf{p}_{\mathbf{y}b} \\ & - \mathbf{w}_b^H \mathbf{R}_{\mathbf{yb}}^H \mathbf{w}_f + \mathbf{w}_b^H \mathbf{R}_{\mathbf{bb}} \mathbf{w}_b + \mathbf{w}_b^H \mathbf{p}_{\mathbf{b}b} \\ & - \mathbf{p}_{\mathbf{y}b}^H \mathbf{w}_f + \mathbf{p}_{\mathbf{b}b}^H \mathbf{w}_b + \sigma_b^2 \end{aligned} \tag{50}$$

with $\mathbf{R}_{\mathbf{yy}} = \mathrm{E}\{\mathbf{y}(k)\mathbf{y}^H(k)\}$, $\mathbf{R}_{\mathbf{bb}} = \mathrm{E}\{\mathbf{b}(k)\mathbf{b}^H(k)\}$, $\mathbf{R}_{\mathbf{yb}} = \mathrm{E}\{\mathbf{y}(k)\mathbf{b}^H(k)\}$, $\mathbf{p}_{\mathbf{y}b} = \mathrm{E}\{\mathbf{y}(k)b^*(k-\kappa)\}$, and $\mathbf{p}_{\mathbf{b}b} = \mathrm{E}\{\mathbf{b}(k)b^*(k-\kappa)\}$. To obtain the optimum weights with the MMSE criterion, we set the gradient of $\mathrm{E}\{|e_t(k)|^2\}$ with respect to the vectors $\mathbf{w}_f^*$ and $\mathbf{w}_b^*$ to zero. This results in

$$\mathbf{w}_{f,\rm opt} = \left(\mathbf{R}_{\mathbf{yy}} - \frac{1}{\sigma_b^2}\mathbf{R}_{\mathbf{yb}}\mathbf{R}_{\mathbf{yb}}^H\right)^{-1} \mathbf{p}_{\mathbf{y}b} \tag{51}$$

$$\mathbf{w}_{b,\rm opt} = \frac{1}{\sigma_b^2}\mathbf{R}_{\mathbf{yb}}^H \mathbf{w}_{f,\rm opt}. \tag{52}$$

Let $\mathbf{w}_{h,\rm opt} = [\omega_0 \quad \omega_1 \quad \cdots \quad \omega_{\gamma-1}]^T$ represent the convolution of $\mathbf{w}_{h,\rm opt}$ and $\mathbf{w}_{f,\rm opt}$ as $\mathbf{H}\mathbf{w}_{f,\rm opt}$, where $\mathbf{H}$ is an $(\alpha+\gamma-1)\times\alpha$ matrix as

$$\mathbf{H} = \begin{bmatrix} \omega_0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ \omega_1 & \omega_0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & & \ddots & & & & & \vdots \\ \omega_{\gamma-1} & \omega_{\gamma-2} & \cdots & \omega_0 & 0 & \cdots & \cdots & 0 \\ 0 & \omega_{\gamma-1} & \omega_{\gamma-2} & \cdots & \omega_0 & 0 & \cdots & 0 \\ \vdots & & \ddots & & & \ddots & & \vdots \\ 0 & \cdots & 0 & \omega_{\gamma-1} & \omega_{\gamma-2} & \cdots & \omega_0 & 0 \\ 0 & \cdots & \cdots & 0 & \omega_{\gamma-1} & \omega_{\gamma-2} & \cdots & \omega_0 \\ \vdots & & & & & \ddots & & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & \omega_{\gamma-1} & \omega_{\gamma-2} \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & \omega_{\gamma-1} \end{bmatrix}. \tag{53}$$

We can further partition $\mathbf{H}$ as $\mathbf{H} = [\mathbf{H}_r^T \quad \mathbf{H}_p^T]^T$, where $\mathbf{H}_r$ is of dimension $(\kappa+1)\times\alpha$, and $\mathbf{H}_p$ is of dimension $(\alpha+\gamma-2-\kappa)\times\alpha$. It is not difficult to see that $\mathbf{H}_r\mathbf{w}_{f,\rm opt}$ corresponds to the precursor response of $\mathbf{H}\mathbf{w}_{f,\rm opt}$, whereas $\mathbf{H}_p\mathbf{w}_{f,\rm opt}$ is the postcursor response. With enough degrees of freedom, CCI is mostly suppressed. The DFGSC output can be modeled as

$$y_s(k) = \mathbf{w}_q^H \mathbf{d}(k) + \nu(k) \tag{54}$$

where $\nu(k)$ is a white noise independent of $\mathbf{d}(k)$. With (54) and some manipulations, we can derive

$$\frac{1}{\sigma_b^2}\mathbf{R}_{\mathbf{yb}}^H = \mathbf{H}_p. \tag{55}$$

From (52), we then obtain $\mathbf{w}_{b,\rm opt} = \mathbf{H}_p\mathbf{w}_{f,\rm opt}$. This result can be restated as

$$\mathbf{w}_{b,\rm opt} = \mathrm{post}\{\mathbf{w}_{h,\rm opt} \otimes \mathbf{w}_{f,\rm opt}\} \tag{56}$$

where $\otimes$ denotes the convolution operation, and $\mathrm{post}\{\cdot\}$ denotes the postcursor-taking operation. This result suggests

an adaptation approach for the training-based MMSE-DFE. Let $\mathbf{w}_f(k)$ and $\mathbf{w}_b(k)$ be the feedforward and feedback weight vectors at time instant k. With reference to (56), we can let

$$\mathbf{w}_b(k) = \text{post}\left\{\mathbf{w}_h(k) \otimes \mathbf{w}_f(k)\right\} \tag{57}$$

in which $\mathbf{w}_h(k)$ is the channel estimate at time instant k. If $\mathbf{w}_h(k)$ converges to $\mathbf{w}_{h,\text{opt}}$, $\mathbf{w}_b(k)$ will converge to $\mathbf{w}_{b,\text{opt}}$ too. The difference between this approach and the conventional method lies in that only $\mathbf{w}_f(k)$ is adapted [not both $\mathbf{w}_f(k)$ and $\mathbf{w}_b(k)$]. While this approach may make no difference for a training-based DFE (since the training sequence is available), it will provide significant improvement for the blind DFE. On one hand, consider a blind DFE scenario in which both $\mathbf{w}_f(k)$ and $\mathbf{w}_b(k)$ are adapted. With $y_t(k) = \mathbf{w}_f^H(k)\mathbf{y}(k) - \mathbf{w}_b^H(k)\hat{\mathbf{b}}(k)$, the update equations for the MMA can be written as

$$\mathbf{w}_f(k+1) = \mathbf{w}_f(k) + \mu_f \phi\left(y_t(k)\right)\mathbf{y}(k) \tag{58}$$

$$\mathbf{w}_b(k+1) = \mathbf{w}_b(k) - \mu_b \phi\left(y_t(k)\right)\hat{\mathbf{b}}(k) \tag{59}$$

where $\phi(y_t(k))$ is the same as that given in (22). From (58) and (59), we see that if there is a decision error, the error will immediately reflect to $\hat{\mathbf{b}}(k)$ and then $\phi(y_t(k))$. Note that the adaptation of $\mathbf{w}_f(k)$ only involves erroneous $\phi(y_t(k))$, whereas that of $\mathbf{w}_b(k)$ involves both erroneous $\hat{\mathbf{b}}(k)$ and erroneous $\phi(y_t(k))$. The two error sources in (59) will make $\mathbf{w}_b(k)$ quite sensitive to decision errors. On the other hand, in the proposed method, only $\mathbf{w}_f(k)$ is adapted, as given in (58). Although the effect of decision error will also pass to $e_s(k)$, which will perturb the adaptation of $\mathbf{w}_a(k)$ and $\mathbf{w}_h(k)$, the influence is smaller. This is because the convergence of $\mathbf{w}_a(k)$ and $\mathbf{w}_h(k)$ in the DFGSC is much faster and more stable than that of the blind DFE (which will empirically be shown in Section VI). By using (57) to calculate $\mathbf{w}_b(k)$, the feedback filter of the blind DFE will perform much better. In one word, with the proposed operation, the resultant CBDFE will be less sensitive to decision errors.

In the startup period, decisions may not be trustworthy. Including decisions in the operation of the CBDFE may affect its stability. Unlike the LBDFE's mode switching, we propose a simple mechanism allowing a smooth transition from FIR linear equalization to DFE. We let the feedback filter be multiplied by a time-varying factor $f(k)$, where $f(k) \leq 1$. Initially, k is small, and we set $f(k)$ small such that the weighting of the feedback filter is small. In this stage, the proposed equalizer will behave much like an FIR linear equalizer. As k increases and decisions become more reliable, we increase $f(k)$. Eventually, $f(k)$ approaches one, and the equalizer becomes a full DFE. As known in [18], the convergence of an adaptive algorithm exhibits exponential decay behavior. Thus, a natural choice for $f(k)$ will be

$$f(k) = 1 - e^{-\xi k} \tag{60}$$

where ξ is a design parameter controlling the increase rate of this factor. With the proposed mechanism, the feedback filter is gradually taken into account, and the error propagation effect will be reduced. The CBDFE will ultimately approach the optimum MMSE-DFE.

V. Spatial Multipath and Spatial Signature Mismatch

In this section, we will show that the proposed STE is robust to two unfavorable phenomena, i.e., spatial multipath propagation and spatial signature mismatch, which are frequently encountered in array signal processing. Let us first consider the spatial multipath propagation problem. Multipath propagation may cause ISI coming from different directions, which induces both coherent interference [26] and angular spread [27]. In this case, the conventional GSC with the simple point distortionless constraint tends to cancel the desired signal itself, which is referred to as signal cancellation. In the proposed STE, even with this spatial ISI, the phenomenon of signal cancellation will not exist. This is due to the use of decision feedback. In the spatial multipath environment, the desired signal will leak to the output of the blocking matrix, producing a correlation between the upper and lower paths in the DFGSC. Thus, the optimum solutions for the interference canceling filter $\mathbf{w}_a$ and the channel estimator $\mathbf{w}_h$ are now coupled together. From (29), we have

$$\mathbf{w}_{\mathbf{c},\text{opt}} = \begin{bmatrix} \mathbf{w}_{a,\text{opt}} \\ \mathbf{w}_{h,\text{opt}} \end{bmatrix} = \mathbf{R}_{\mathbf{c}}^{-1} \begin{bmatrix} \mathbf{B}^H\mathbf{R}_{\mathbf{x}} \\ \mathbf{R}_{\mathbf{p}}^H \end{bmatrix} \mathbf{w}_q \tag{61}$$

where

$$\begin{aligned} \mathbf{R}_{\mathbf{p}} &= \text{E}\left\{\mathbf{x}(k)\hat{\mathbf{b}}^H(k)\right\} \\ &= \sigma_b^2 \left[\sum_{m=0}^{M_0-1}\mathbf{h}_m(0) \sum_{m=0}^{M_0-1}\mathbf{h}_m(1) \cdots \sum_{m=0}^{M_0-1}\mathbf{h}_m(\gamma-1)\right] \end{aligned} \tag{62}$$

$$\mathbf{R}_{\mathbf{c}} = \begin{bmatrix} \mathbf{B}^H\mathbf{R}_{\mathbf{x}}\mathbf{B} & \mathbf{M} \\ \mathbf{M}^H & \sigma_b^2\mathbf{I}_\gamma \end{bmatrix} \tag{63}$$

with

$$\mathbf{M} = \sigma_b^2\mathbf{B}^H \left[\sum_{m=0}^{M_0-1}\mathbf{h}_m(0) \sum_{m=0}^{M_0-1}\mathbf{h}_m(1) \cdots \sum_{m=0}^{M_0-1}\mathbf{h}_m(\gamma-1)\right] \tag{64}$$

which is the correlation matrix between the blocking matrix output $\mathbf{B}^H\mathbf{x}(k)$ and the decision vector $\hat{\mathbf{b}}(k)$. Using the inversion identity for subblock matrices [28], we can find the inverse of $\mathbf{R}_{\mathbf{c}}$, and so, $\mathbf{w}_{\mathbf{c},\text{opt}}$ in (61), i.e.,

$$\begin{aligned} \mathbf{w}_{\mathbf{c},\text{opt}} = \Bigg[& (\mathbf{B}^H\mathbf{R}_{\mathbf{z}}\mathbf{B})^{-1}\mathbf{B}^H\mathbf{R}_{\mathbf{z}}\mathbf{w}_q \\ & \left[\sum_{m=0}^{M_0-1}\mathbf{h}_m(0) \cdots \sum_{m=0}^{M_0-1}\mathbf{h}_m(\gamma-1)\right]^H \\ & \times \left(\mathbf{I} - \mathbf{B}(\mathbf{B}^H\mathbf{R}_{\mathbf{z}}\mathbf{B})^{-1}\mathbf{B}^H\mathbf{R}_{\mathbf{z}}\right)\mathbf{w}_q \Bigg] \end{aligned} \tag{65}$$

and thus the coupled $\mathbf{w}_{a,\mathrm{opt}}$ and $\mathbf{w}_{h,\mathrm{opt}}$ can be written as

$$\mathbf{w}_{a,\mathrm{opt}} = (\mathbf{B}^H\mathbf{R}_\mathbf{z}\mathbf{B})^{-1}\mathbf{B}^H\mathbf{R}_\mathbf{z}\mathbf{w}_q \tag{66}$$

$$\mathbf{w}_{h,\mathrm{opt}} = \left[\sum_{m=0}^{M_0-1}\mathbf{h}_m(0)\cdots\sum_{m=0}^{M_0-1}\mathbf{h}_m(\gamma-1)\right]^H \times(\mathbf{w}_q - \mathbf{B}\mathbf{w}_{a,\mathrm{opt}}). \tag{67}$$

The minimum J of the DFGSC, as that given in (35), can then be solved to be

$$\begin{aligned} J_{\min} &= \mathbf{w}_q^H\mathbf{R}_\mathbf{x}\mathbf{w}_q - \mathbf{w}_q^H[\,\mathbf{R}_\mathbf{x}\mathbf{B}\quad\mathbf{R}_\mathbf{p}\,]\mathbf{R}_\mathbf{c}^{-1}\begin{bmatrix}\mathbf{B}^H\mathbf{R}_\mathbf{x}\\ \mathbf{R}_\mathbf{p}^H\end{bmatrix}\mathbf{w}_q \\ &= \mathbf{w}_q^H\mathbf{R}_\mathbf{z}\mathbf{w}_q - \mathbf{w}_q^H\mathbf{R}_\mathbf{z}\mathbf{B}(\mathbf{B}^H\mathbf{R}_\mathbf{z}\mathbf{B})^{-1}\mathbf{B}^H\mathbf{R}_\mathbf{z}\mathbf{w}_q \\ &= \mathbf{w}_q^H\mathbf{R}_\mathbf{z}\mathbf{w}_{\mathrm{opt}}. \end{aligned} \tag{68}$$

From (66)–(68), we see that while $\mathbf{w}_{a,\mathrm{opt}}$ and $\mathbf{w}_{h,\mathrm{opt}}$ are changed, $J_{\min}$ (and so the error signal) still contains no desired signal. The leaky desired signal in the lower path of the DFGSC is not correlated with the error signal. When optimizing $\mathbf{w}_a$, the output desired signal power is not minimized, and so, no signal cancellation occurs.

We then consider the spatial signature mismatch problem. If the knowledge of the desired signal's main DOA is erroneous, $\mathbf{w}_q$ no longer matches the desired signal's spatial signature, and $\mathbf{B}$ cannot obstruct the desired signal. In this case, signal cancellation may occur as well [29]. In the case of spatial multipath propagation, we have already proven that for the DFGSC, even if there is leaky desired signal in the output of the blocking matrix, $\mathbf{w}_a$ will not cancel the desired signal. The spatial signature mismatch can be considered a signal-leaking problem similar to that in spatial multipath propagation. Thus, no signal cancellation will occur in the DFGSC. The derivation details, however, are omitted.

For these scenarios, the expression of the optimum output SINR is the same as that given in (47), but now, $P_\mathbf{d}$ and $P_{o,\min}$ are changed. The output desired signal power becomes

$$P_\mathbf{d} = \sigma_b^2\left|(\mathbf{w}_q - \mathbf{B}\mathbf{w}_{a,\mathrm{opt}})^H \times\left[\sum_{m=0}^{M_0-1}\mathbf{h}_m(0)\sum_{m=0}^{M_0-1}\mathbf{h}_m(1)\cdots\sum_{m=0}^{M_0-1}\mathbf{h}_m(\gamma-1)\right]\right|^2 \tag{69}$$

which is different from (33). In the calculation of $P_{o,\min}$, (15) should be used instead of (16). Moreover, due to the existence of the desired signal in the input of the interference canceling filter, the excess MSE yielded by the LMS algorithm will be different too. Let $J_{\mathrm{ex}}^\mathbf{d}(\infty)$ denote the excess MSE value for the desired signal in these leaky scenarios. Using (44), we can have

$$J_{\mathrm{ex}}^\mathbf{d}(\infty) = J_{\min}\frac{\mu_a\lambda_1(\mathbf{B}^H\mathbf{R}_\mathbf{d}\mathbf{B})}{2-\mu_a\lambda_1(\mathbf{B}^H\mathbf{R}_\mathbf{d}\mathbf{B})}. \tag{70}$$

The result in (70) is due to the fact that only one eigenvalue of $\mathbf{B}^H\mathbf{R}_\mathbf{d}\mathbf{B}$ is nonzero. The expression of the steady-state output SINR for the GSC structure with spatial ISI or spatial signature mismatch is then slightly modified to

$$\mathrm{SINR}_{\mathrm{LMS}} = \frac{P_\mathbf{d} + J_{\mathrm{ex}}^\mathbf{d}(\infty)}{P_{o,\min} + J_{\mathrm{ex}}(\infty) - (P_\mathbf{d} + J_{\mathrm{ex}}^\mathbf{d}(\infty))} \tag{71}$$

in which $P_\mathbf{d}$ and $P_{o,\min}$ should also be changed to use (69) and (15), respectively.

TABLE I
PARAMETERS USED IN SIMULATIONS OF PART A: (a) FILTER LENGTH AND (b) STEP SIZES

	$\mathbf{w}_a$	$\mathbf{w}_f$	$\mathbf{w}_b$	$\mathbf{w}_h$
All schemes	4	7	7	3

(a)

	$\mathbf{w}_\mathrm{a}$	$\mathbf{w}_\mathrm{f}$	$\mathbf{w}_\mathrm{b}$
DFGSC-CBDFE (s & t)	5×10^{-4}	1×10^{-2}	/
GSC-LBDFE (s)	5×10^{-4}	1×10^{-4}	2×10^{-4}
GSC-LBDFE (t)	5×10^{-4}	1×10^{-2}	1×10^{-2}
DFGSC-LBDFE (s)	5×10^{-4}	1×10^{-4}	2×10^{-4}
DFGSC-LBDFE (t)	5×10^{-4}	1×10^{-2}	1×10^{-2}

s: start-up, t: tracking

(b)

VI. SIMULATIONS

Computer simulations are conducted to demonstrate the effectiveness of the proposed STE (called DFGSC-CBDFE in this section for clarity) and verify our analytic results. For comparison, we also consider the hybrid of the conventional GSC and LBDFE (called GSC-LBDFE hereafter) and the hybrid of the DFGSC and LBDFE (called DFGSC-LBDFE hereafter). Note that the latter scheme is used for the comparison of CBDFE and LBDFE. In all cases, we assume a ULA with five omnidirectional antennas spaced half a wavelength apart. The main DOA of the desired signal is known *a priori*. Only the point distortionless constraint to the main DOA is used for the GSC in all three schemes. The received desired signal is corrupted by CCI, ISI, and AWGN. In the first part, we only consider channels with temporal ISI. In the second part, we consider more realistic channels where both temporal and spatial ISIs are present. In all figures, at least 500 simulation runs are averaged to obtain each simulated result.

A. Channels With Temporal ISI Only

In this set of simulations, we consider channels with temporal ISI only. We compare the performance of DFGSC-CBDFE with GSC-LBDFE and DFGSC-LBDFE. The transmitted symbols are randomly generated and then modulated by quadrature phase-shift keying. The CCI-to-noise ratio (CCINR) is set as 25 dB, and the signal-to-noise ratio (SNR) is 15 dB. The channel for the desired signal coming from 0° is [0.407 0.815 0.407] [1, p. 616], and the channel for CCI coming from -60° is [1 0 0]. The parameters used for the adaptive GSC, DFGSC, LBDFE, and CBDFE are listed in Table I. The step sizes for $\mathbf{w}_h$ in those decision feedback schemes are 2×10^{-3}. The decision-directed MMSE training starts after 500 iterations. As mentioned, the factor $f(k)$ for the CBDFE is used to reduce error propagation. In this case, ξ is chosen to be 0.01, and $f(k)$ approaches

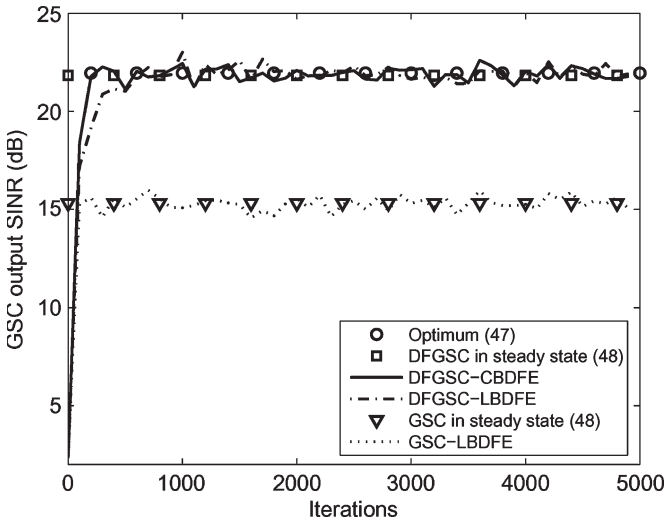


Fig. 3. Learning curves of GSC output SINR for different schemes in suppressing CCI.

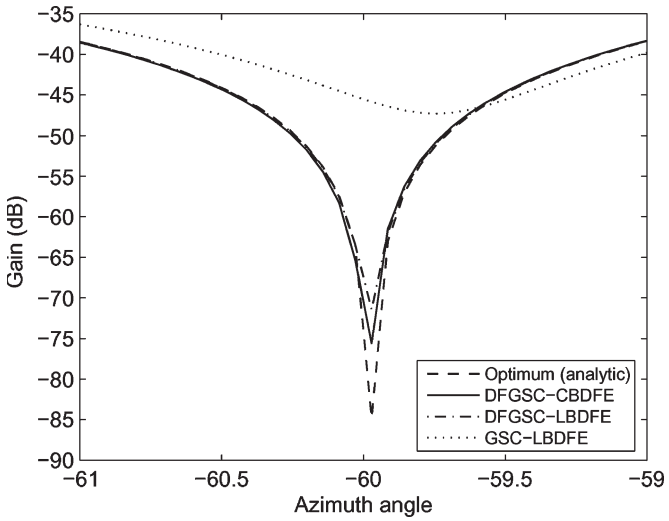


Fig. 4. Beam patterns (enlarged region around CCI's DOA) of different schemes in Fig. 3 after 5000 iterations.

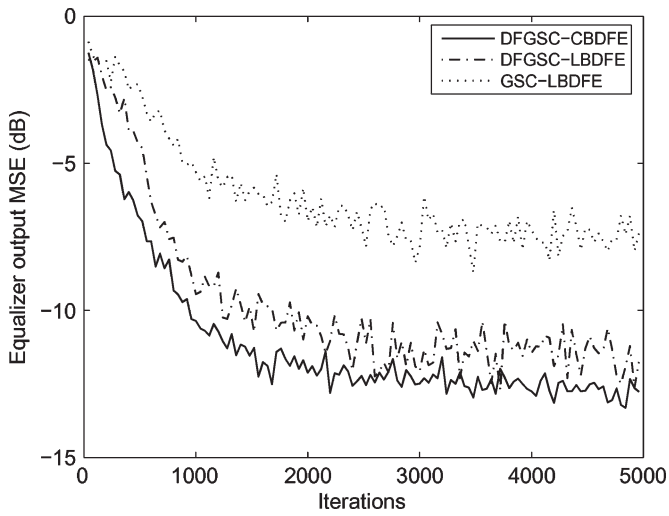


Fig. 5. Learning curves of equalizer output MSE for different schemes.

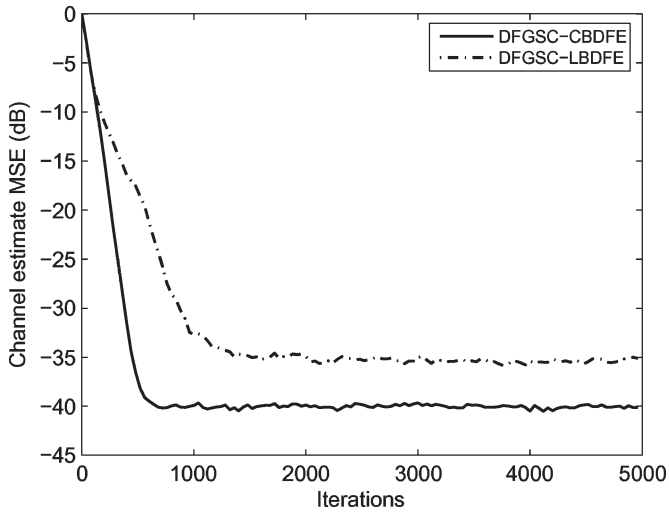


Fig. 6. Learning curves of channel estimate MSE for different schemes.

unity at around 500 iterations. Fig. 3 shows the GSC output SINR for DFGSC-CBDFE, GSC-LBDFE, and DFGSC-LBDFE. In addition, the optimum SINR and the steady-state SINR obtained with the LMS algorithm are shown. From Fig. 3, we see that all schemes are comparable in convergence rate, but the SINR achieved by those schemes with the adaptive DFGSC is higher than that with the conventional adaptive GSC. As expected, the adaptive DFGSC can approach the optimum solution much more closely, which means that the effect of the excess MSE induced by the LMS algorithm is smaller. In terms of GSC output SINR and convergence rate, DFGSC-CBDFE provides the best performance. Fig. 4 reveals the beam patterns (enlarging the region around the CCI's DOA) of those schemes used in Fig. 3 after 5000 iterations. The improvement of the decision feedback operation can clearly be seen as well. Fig. 5 shows the equalizer output MSE, i.e., $\mathrm{E}\{|b(k-\kappa)-y_t(k)|^2\}$, for all three schemes. From the figure, we see that both schemes with adaptive DFGSC work well in suppressing the remaining ISI. However, DFGSC-CBDFE performs better than DFGSC-LBDFE. This means that the CBDFE is more effective in ISI suppression. In Fig. 6, we show the MSE of the channel estimate, i.e., $\mathrm{E}\{|[\mathbf{h}_0(0) \quad \mathbf{h}_0(1) \quad \cdots \quad \mathbf{h}_0(\gamma-1)]^H\mathbf{w}_q-\mathbf{w}_h(k)|^2\}$, for the schemes with the adaptive DFGSC. We observe that these MSE values become small when the decision is reliable (about -40 dB for DFGSC-CBDFE and about -35 dB for DFGSC-LBDFE). With the accurate channel estimate, the DFGSC can eliminate the desired signal from the error signal, and the CBDFE can construct a better feedback filter for equalization. This is also where the improvement comes from.

To show the robustness of the DFGSC against spatial signature mismatch, we repeat the aforementioned scenario with a 5°, DOA mismatch of the desired signal, i.e., the actual main DOA is -5°, and the presumed main DOA is 0°. We plot the resultant GSC output SINR, beam patterns, and equalizer output MSE in Figs. 7–9, respectively. In Fig. 7, we observe that the SINR performance gap between DFGSC-CBDFE and DFGSC-LBDFE becomes larger, and GSC-LBDFE completely fails due to signal cancellation. In addition, Fig. 8 shows that DFGSC-CBDFE can keep the beam pattern very close to the

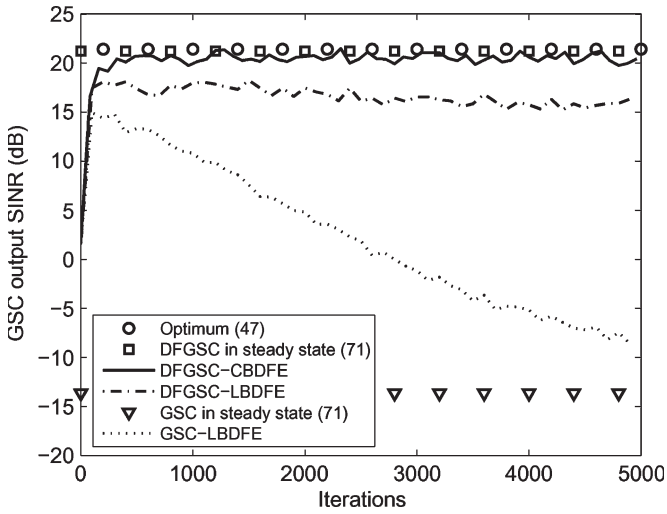


Fig. 7. Learning curves of GSC output SINR for different schemes in which the actual main DOA is $-5°$, and the presumed main DOA is $0°$.

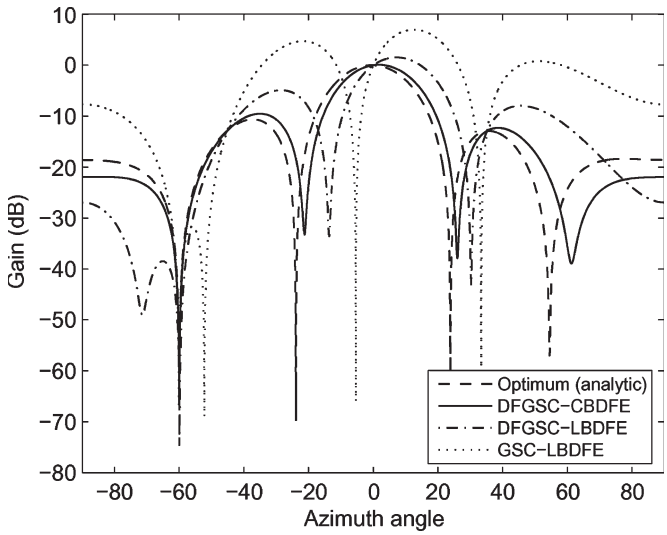


Fig. 8. Beam patterns of different schemes in Fig. 7 after 5000 iterations in which the actual main DOA is $-5°$, and the presumed main DOA is $0°$.

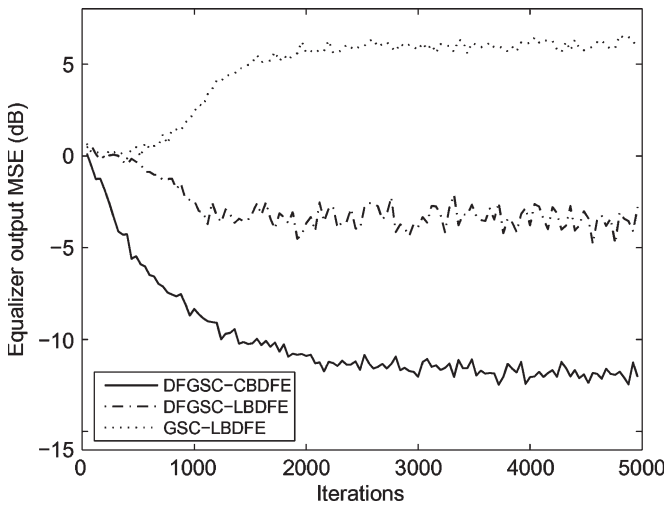


Fig. 9. Learning curves of equalizer output MSE for different schemes in which the actual main DOA is $-5°$, and the presumed main DOA is $0°$.

TABLE II
PARAMETERS USED IN SIMULATIONS OF PART B:
(a) STEP SIZES AND (b) CHANNEL SETTINGS

	$\mathbf{w}_a$	$\mathbf{w}_f$	$\mathbf{w}_b$
DFGSC-CBDFE (s)	5×10^{-4}	2×10^{-5}	/
DFGSC-CBDFE (t)	5×10^{-4}	3×10^{-3}	/
GSC-LBDFE (s)	5×10^{-4}	1×10^{-5}	2×10^{-5}
GSC-LBDFE (t)	5×10^{-4}	3×10^{-3}	3×10^{-3}
DFGSC-LBDFE (s)	5×10^{-4}	1×10^{-5}	2×10^{-5}
DFGSC-LBDFE (t)	5×10^{-4}	3×10^{-3}	3×10^{-3}

s: start-up, t: tracking

(a)

	DOA	Channel
Desired signal	$0°$	[0.1 1.0 0.2]
Spatial ISI	$2°$	[0.1 0.5 0.1]
	$30°$	[0.0 0.2 0.7]
	$50°$	[0.2 0.6 0.0]
CCI	$-60°$	[1.0 0.0 0.0]

(b)

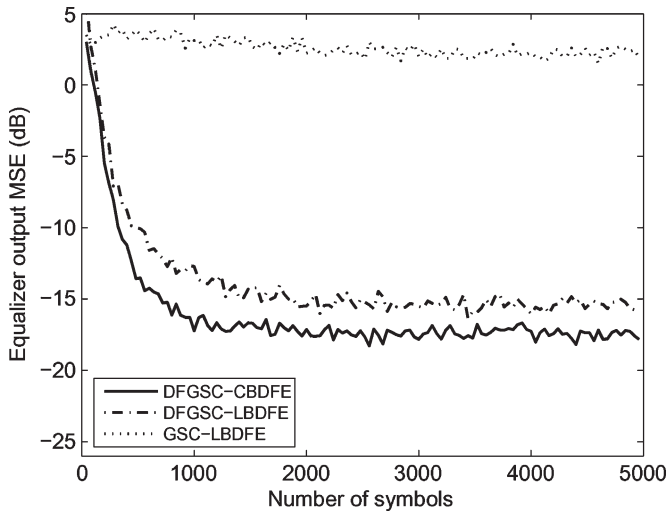


Fig. 10. Learning curves of equalizer output MSE for different schemes with spatial ISI.

optimum. In Fig. 9, we see that DFGSC-CBDFE gives the smallest equalizer output MSE and, thus, performs best. All these indicate the DFGSC is better than the conventional GSC, and the CBDFE is better than the LBDFE.

B. Channels With Both Temporal and Spatial ISI

In this section, we consider a more general case where ISI comes from different directions and time instants. Again, the CCINR is fixed at 25 dB, and the SNR is 20 dB. The filter length for the blind DFE is the same as given previously. The step sizes and channel settings used are shown in Table II. From the table, we see that the ISI comes from different directions, and the effect of both coherent interference and angular spread is included. Furthermore, the step sizes for $\mathbf{w}_h$ are 1×10^{-3}. The parameter ξ is selected as 0.005, for which $f(k)$ is close to unity around 1000 iterations. Fig. 10 shows the equalizer output MSE with 16 quadrature amplitude modulation for the three schemes. To observe the convergence behavior of these schemes, we exclude the decision-directed mode and use the MMA blind algorithm all the way first. We see that DFGSC-CBDFE outperforms the other two schemes in terms of both convergence rate and equalizer output MSE. The channel

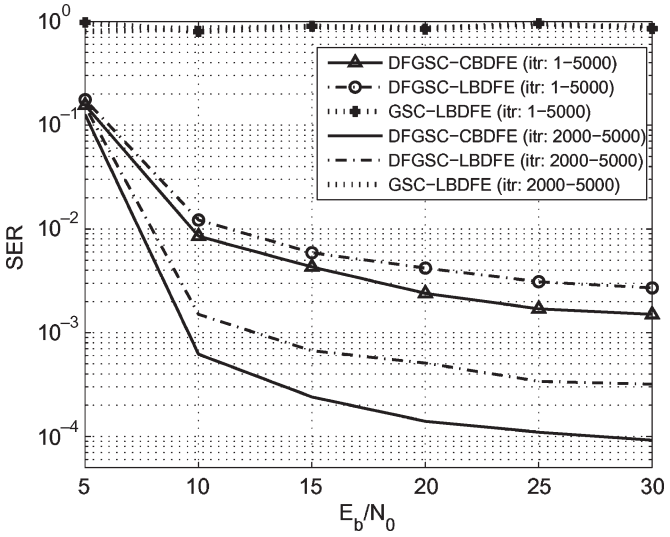


Fig. 11. SER performance for the scenario in Fig. 10.

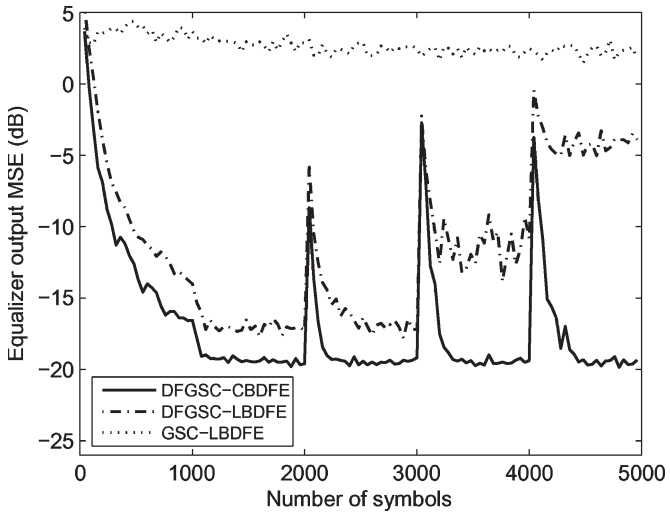


Fig. 12. Learning curves of equalizer output MSE for different schemes with spatial ISI (artificial errors added as five consecutive errors from the 2000th iteration, ten interleaved errors from the 3000th iteration, and ten consecutive errors from the 4000th iteration).

estimator in the proposed scheme really helps to construct a more stable and efficient blind DFE. Then, the symbol-error-rate (SER) performance for the three schemes under the same settings is given in Fig. 11. The result is observed under different iteration intervals. Again, DFGSC-CBDFE performs best and offers smaller SER in all cases. Note that GSC-LBDFE fails as well due to signal cancellation.

Finally, we repeat the same experiment with the decision-directed MMSE training in the tracking period (after 1000 iterations). However, some artificial errors are added in the decision process. To force an error to occur, we randomly shift the decision to a constellation point near its true value. The errors are added as follows: five consecutive errors from the 2000th iteration, ten interleaved errors from the 3000th iteration, and ten consecutive errors from the 4000th iteration. The result is presented in Fig. 12. Again, it shows that DFGSC-CBDFE performs better than DFGSC-LBDFE. The proposed DFGSC-CBDFE can reconverge after the bursty errors, but DFGSC-LBDFE diverges. This clearly shows that the CBDFE is less sensitive to decision errors and makes the whole processing scheme more reliable. We then conclude that DFGSC-CBDFE is the best hybrid scheme for the scenario we consider.

VII. Conclusion

In this paper, a new adaptive STE has been developed for the suppression of both CCI and ISI. The proposed STE introduces a hybrid of an adaptive DFGSC and an adaptive CBDFE. With the main DOA known *a priori*, training sequences are not required for the adaptation of the whole STE. We show that the included channel estimator can not only improve the performance of the conventional adaptive GSC but also make the blind DFE more reliable. For spatial processing, the DFGSC improves the CCI suppression capability when implemented with the LMS algorithm. In addition, the adaptive DFGSC with the simple point distortionless constraint is robust against multipath propagation environments and spatial signature errors. For temporal processing, the proposed CBDFE can have better performance than the LBDFE. With our special adaptation, the problem of error propagation is reduced. Simulation results verified our analysis and confirmed that the proposed STE can achieve good performance, even in severe channel environments.

References

[1] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.

[2] R. B. Ertel, P. Cardieri, K. W. Sowerby, T. S. Rappaport, and J. H. Reed, "Overview of spatial channel models for antenna array communication systems," *IEEE Pers. Commun.*, vol. 5, no. 1, pp. 10–22, Feb. 1998.

[3] R. A. Monzingo and T. W. Miller, *Introduction to Adaptive Arrays*. New York: Wiley, 1980.

[4] L. C. Godara, "Application of antenna arrays to mobile communications. II. Beam-forming and direction-of-arrival considerations," *Proc. IEEE*, vol. 85, no. 8, pp. 1195–1245, Aug. 1997.

[5] G. V. Tsoulos, Ed., *Adaptive Antennas for Wireless Communications*, Piscataway, NJ: IEEE Press, 2001.

[6] R. Kohno, "Spatial and temporal communication theory using adaptive antenna array," *IEEE Pers. Commun.*, vol. 5, no. 1, pp. 28–35, Feb. 1998.

[7] A. J. Paulraj and B. C. Ng, "Space–time modems for wireless personal communications," *IEEE Pers. Commun.*, vol. 5, no. 1, pp. 36–48, Feb. 1998.

[8] J.-W. Liang, J.-T. Chen, and A. J. Paulraj, "A two-stage hybrid approach for CCI/ISI reduction with space–time processing," *IEEE Commun. Lett.*, vol. 1, no. 6, pp. 163–165, Nov. 1997.

[9] N. Ishii and R. Kohno, "Spatial and temporal equalization based on an adaptive tapped delay line array antenna," *IEICE Trans. Commun.*, vol. E78-B, no. 8, pp. 1162–1169, Aug. 1995.

[10] M. Chiani and A. Zanella, "Spatial and temporal equalization for broadband wireless indoor networks at millimeter waves," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 10, pp. 1725–1734, Oct. 1999.

[11] M.-L. Leou, C.-C. Yeh, and H.-J. Li, "A novel hybrid of adaptive array and equalizer for mobile communications," *IEEE Trans. Veh. Technol.*, vol. 49, no. 1, pp. 1–10, Jan. 2000.

[12] K. Hayashi and S. Hara, "A new spatio-temporal equalization method based on estimated channel response," *IEEE Trans. Veh. Technol.*, vol. 50, no. 5, pp. 1250–1259, Sep. 2001.

[13] G. V. Tsoulos, "Smart antennas for mobile communication systems: Benefits and challenges," *Electron. Commun. Eng. J.*, vol. 11, no. 2, pp. 84–94, Apr. 1999.

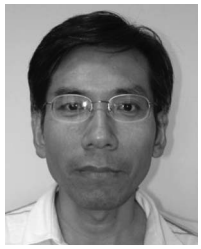
[14] J. C. Liberti and T. S. Rappaport, *Smart Antennas for Wireless Communications: IS-95 and Third Generation CDMA Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1999.

[15] B. Suard, G. Xu, H. Liu, and T. Kailath, "Uplink channel capacity of space-division-multiple-access schemes," *IEEE Trans. Inf. Theory*, vol. 44, no. 4, pp. 1468–1476, Jul. 1998.

[16] G. M. Galvan-Tejada and J. G. Gardiner, "Theoretical model to determine the blocking probability for SDMA systems," *IEEE Trans. Veh. Technol.*, vol. 50, no. 5, pp. 1279–1288, Sep. 2001.
[17] L. J. Griffiths and C. W. Jim, "Alternative approach to linear constrained adaptive beamforming," *IEEE Trans. Antennas Propag.*, vol. AP-30, no. 1, pp. 27–34, Jan. 1982.
[18] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Upper Saddle River, NJ: Prentice-Hall, 1996.
[19] J. Labat, O. Macchi, C. Laot, "Adaptive decision feedback equalization: Can you skip the training period?" *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 921–930, Jul. 1998.
[20] Y. Lee and W.-R. Wu, "A robust adaptive generalized sidelobe canceller with decision feedback," *IEEE Trans. Antennas Propag.*, vol. 53, no. 11, pp. 3822–3832, Nov. 2005.
[21] O. L. Frost, III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.
[22] J. Yang, J. J. Werner, and G. A. Dumont, "The multimodulus blind equalization and its generalized algorithms," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 5, pp. 997–1015, Jun. 2002.
[23] L. M. Garth, "A dynamic convergence analysis of blind equalization algorithms," *IEEE Trans. Commun.*, vol. 49, no. 4, pp. 624–634, Apr. 2001.
[24] P. K. Shukla and L. F. Turner, "Channel-estimation-based adaptive DFE for fading multipath radio channels," *Proc. Inst. Electr. Eng. I*, vol. 138, no. 6, pp. 525–543, Dec. 1991.
[25] A. A. Rontogiannis and K. Berberidis, "Efficient decision feedback equalization for sparse wireless channels," *IEEE Trans. Wireless Commun.*, vol. 2, no. 3, pp. 570–581, May 2003.
[26] T.-J. Shan and T. Kailath, "Adaptive beamforming for coherent signals and interference," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 3, pp. 527–536, Jun. 1985.
[27] S. Valaee, B. Champagne, and P. Kabal, "Parametric localization of distributed sources," *IEEE Trans. Signal Process.*, vol. 43, no. 9, pp. 2144–2153, Sep. 1995.
[28] G. H. Golub and C. F. Van Loan, *Matrix Computation*, 3rd ed. Baltimore, MD: The Johns Hopkins Univ. Press, 1996.
[29] B. Widrow, K. Duvall, R. Gooch, and W. Newman, "Signal cancellation phenomena in adaptive antennas: Causes and cures," *IEEE Trans. Antennas Propag.*, vol. AP-30, no. 3, pp. 469–478, May 1982.

Yinman Lee (M'06) was born in Hong Kong. He received the Ph.D. degree from National Chiao Tung University, Hsinchu, Taiwan, R.O.C., in 2006.

He is currently with the Graduate Institute of Communication Engineering, National Chi Nan University, Nantou, Taiwan. His research interests include adaptive signal processing, wireless communications, and multiple antenna systems.



Wen-Rong Wu (M'89) received the B.S. degree in mechanical engineering from Tatung Institute of Technology, Taipei, Taiwan, R.O.C., in 1980, and the M.S. degrees in mechanical and electrical engineering and the Ph.D. degree in electrical engineering from the State University of New York, Buffalo, in 1985, 1986, and 1989, respectively.

Since August 1989, he has been a faculty member with the Department of Communication Engineering, National Chiao Tung University, Hsinchu, Taiwan. His research interests include statistical signal processing and digital communications.