On Channel Estimation and Detection for Multicarrier Signals in Fast and Selective Rayleigh Fading Channels

Yang-Seok Choi, *Member, IEEE*, Peter J. Voltz, *Member, IEEE*, and Frank A. Cassara, *Senior Member, IEEE*

***Abstract*—Time-domain channel estimation and detection techniques are presented for multicarrier signals in a fast and frequency-selective Rayleigh fading channel. As a consequence of the time-varying channel, the orthogonality between subcarriers is destroyed in conventional frequency-domain approaches, resulting in interchannel interference, which increases an irreducible error floor in proportion to the normalized Doppler frequency. An important feature of the proposed technique is the ability to exploit the time-selective channel as a provider of time diversity. This enables us to achieve performance superior to any other structure without increasing bandwidth or incorporating redundancy. In order to reduce the complexity of the estimator, we apply the theory of optimal low rank approximation to a minimum mean squared error channel estimator and present theoretical calculation of mean squared error and simulations to confirm that the estimator is robust to changes in channel characteristics.**

***Index Terms*—Channel estimation, fast fading, multicarrier signals, OFDM, time diversity, time-varying channel.**

I. Introduction

THE DEMAND for high rate data transmission increases rapidly. To meet this demand, some straightforward possibilities include reducing the symbol duration or using higher order modulation techniques. In the former method, the signals received through the multipath channel suffer from severe intersymbol interference (ISI) since the delay spread becomes much larger than the symbol duration. To correctly detect the transmitted data, a complex equalizer is required. In DS-CDMA, the structure of the rake receiver becomes complicated due to the increased number of RAKE fingers for a fixed delay resolution. Generally, the complexity of a channel equalizer is proportional to the delay spread while the complexity of maximum likelihood sequence estimator (MLSE), which is known as one of the best equalization schemes in the sense of minimizing bit error rate (BER), increases exponentially with the delay spread. By using higher order modulations, faster data transmissions can be achieved at the expense of higher transmitted power. However, in DS-CDMA, it unavoidably increases the required E_b/I_o for reliable data communications, which means overall capacity cannot increase.

In order to combat the multipath, the symbol duration must be significantly larger than the channel delay spread. In orthogonal frequency division multiplexing (OFDM) [1]–[4], the entire channel is divided into many narrow subchannels. Splitting the high-rate serial data stream into many low-rate parallel streams, each parallel stream modulates orthogonal subcarriers by means of the inverse fast Fourier transform (IFFT). If the bandwidth of each subcarrier is much less than the channel coherence bandwidth, a frequency flat channel model can be assumed for each subcarrier. Moreover, inserting a cyclic prefix (or guard interval) results in an inter-symbol interference (ISI) free channel assuming that the length of the guard interval is greater than the delay spread of the channel. Therefore, the effect of the multipath channel on each subcarrier can be represented by a single complex multiplier, affecting the amplitude and phase of each subcarrier. Hence, the equalizer at the receiver can be implemented by a set of complex multipliers, one for each subcarrier.

Furthermore, in multicarrier CDMA (MC-CDMA) [5], by spreading parallel data in the frequency domain using an orthogonal code such as the Walsh–Hadamard code, frequency diversity can be obtained in a frequency-selective channel without additional redundancy or increasing the bandwidth since each data bit is distributed over the entire frequency band. The MC-CDMA technique is capable not only of mitigating the ISI, but of exploiting the multipath as well. With proper detection techniques [5], [6], it is shown to suffer only slightly from inter-user interference, whereas DS-CDMA typically experiences significantly higher inter-user interference. This property enables the use of higher order modulation for higher bandwidth efficiency. The multicarrier transmission techniques are among the most promising data transmission schemes in wired and wireless communications.

Despite these advantages, however, the increased symbol duration causes two adverse effects in a time-varying channel. The change in the channel from symbol to symbol is more significant than the single carrier transmission system. This makes channel estimation a difficult task. Moreover, time variations of the channel within a multicarrier symbol lead to a loss of subchannel orthogonality, resulting in interchannel interference (ICI) and leading to an irreducible error floor in conventional receivers. In [7], the minimum mean squared error (MMSE) channel estimator has been proposed using both time- and frequency-domain correlation functions. This can effectively de-

Paper approved by A. Ahlen, the Editor for Modulation & Signal Design of the IEEE Communications Society. Manuscript received April 12, 2000; revised October 16, 2000, and January 17, 2001. This work was supported by a research contract from Fujitsu Laboratories, Ltd, Kawasaki, Japan to Polytechnic University, Brooklyn, N.Y.

Y.-S. Choi was with National Semiconductor, East Brunswick, NJ 08816 USA. He is now with AT&T Laboratories–Research, Middletown, NJ 07748 USA.

P. J. Voltz is with LayerOne Wireless Technology, Melville, NY11747 USA.

F. A. Cassara is with the Department of Electrical Engineering, Polytechnic University, Farmingdale, NY 11735 USA.

Publisher Item Identifier S 0090-6778(01)06941-0.

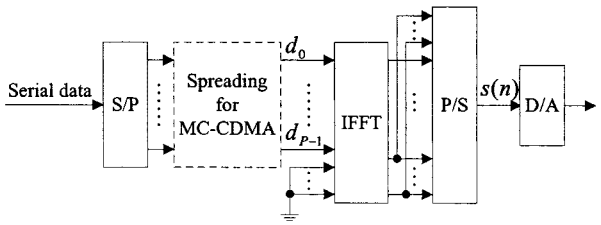


Fig. 1. Transmitter structure of multicarrier systems.

crease the number of pilot symbols to be used for the estimation. By exploiting a separation property, the singular value decomposition greatly reduces the complexity [7], [8]. This frequency-domain approach suffers from ICI, however, since there was no attempt to cancel it. With regard to ICI, a frequency-domain equalizer is presented in [9] under the assumption that the channel impulse response varies with time in a linear fashion. Nevertheless, in rapidly time-varying channel, the assumption does not hold.

In mobile communications, the time-varying channel is a main obstacle to data detection since it destroys the orthogonality of the multicarrier signals. Surprisingly, however, we will show that the time-varying nature of the channel can be exploited as a provider of time diversity provided that a proper detection technique is adopted. The distinct aspect here from previous works is that we make full use of the time-selective channel, and channel estimation and equalization are performed in the time domain. Hence, we can improve the BER performance without generating ICI. In the following, we apply the successive detection method and demonstrate its superior performance. Theoretical calculations of the mean squared error (MSE) together with simulations show the time-domain estimator, even when using a low-rank approximation, is a robust estimator.

The paper is organized as follows. In Section II, the channel and system model for the multicarrier signals are described and an analysis of the ICI is presented. When the impulse response of the channel is ideally known, several detection techniques are addressed together with performance comparisons in Section III. The time-domain MMSE channel estimator is presented in Section IV. Finally, conclusions are drawn in Section V.

II. System Model and Conventional Detection

A. System Model

In multicarrier systems, the symbol duration is increased by splitting the high-rate serial data stream into many low-rate parallel streams. As shown in Fig. 1, the same basic structure is used in both OFDM and MC-CDMA. The difference is that in OFDM an individual data symbol is carried on a single subcarrier, whereas in MC-CDMA each individual data symbol is spread across the entire set of subcarriers by means of the Walsh–Hadamard codes. For OFDM, the data vector, $\boldsymbol{d} = [d_0\ d_1 \dots d_{P-1}]^{\mathrm{T}}$, consists of individual data bits, whereas for MC-CDMA it takes the form $\boldsymbol{d} = \widetilde{d}_0\boldsymbol{c}_0 + \widetilde{d}_1\boldsymbol{c}_1 + \cdots + \widetilde{d}_{P-1}\boldsymbol{c}_{P-1}$, where the $\widetilde{d}_i$ are the data bits and the $\boldsymbol{c}_i$ are P-length Walsh–Hadamard codes.

Throughout this paper, we focus on the OFDM structure. In order to eliminate interference between parallel data streams, each low-rate data stream modulates orthogonal subcarriers by means of the IFFT as shown in Fig. 1. A cyclic prefix is then added to eliminate the effect of the ISI. For proper digital-to-analog (D/A) conversion and lowpass filtering (LPF), we should include unused subcarriers (virtual subcarriers) which are contiguous with the occupied subcarriers. The number of data bearing subcarriers is called the number of active subcarriers. Without loss of generality, assume that the active subcarriers are those with indices 0 to $P-1$. Then, in discrete time, the transmitted multicarrier signal with a number, P, of active subcarriers can be written as follows:

$$s(n) = \sqrt{\frac{E_s}{N}} \sum_{k=0}^{P-1} d_k \mathrm{e}^{j2\pi nk/N}, \quad -L \le n \le N-1 \qquad (1)$$

where E_s is the symbol energy per subcarrier, N is the FFT size, L is the length of the guard interval, and

$$E\left\{|d_k|^2\right\} = 1, \quad E\left\{|s(n)|^2\right\} = \frac{P}{N}E_s.$$

By the central limit theorem, the transmitted signal $s(n)$ can be modeled as a colored complex Gaussian process with zero mean provided P is sufficiently large.

B. Channel Model

In many radio channels, there may be more than one path from transmitter to receiver. Such multiple paths (multipath) may be due to atmospheric reflection, refraction, or reflections from buildings and other objects. The time delays and attenuation factors of the different paths are generally time-varying in mobile communications. If we assume the well-known wide sense stationary uncorrelated scattering (WSSUS) model [10], the channel is characterized by its delay power spectrum (or multipath intensity profile) and scattering function. If the signal is band-limited, then the time-varying diffuse multipath channel can be represented as a tapped delay line with time-varying coefficients and fixed tap spacing. In this tapped delay line channel model, the length of the tapped delay line is determined by the duration of the delay power spectrum, or delay spread (T_d), and the tap spacing must be equal to or less than the reciprocal of the passband bandwidth [11]. The delay power spectrum also determines the power distribution among the taps. The scattering function describes time-varying behavior of each tap. It is determined by the Doppler frequency and antenna structure. In this paper, we assume that the multipath intensity profile has an exponential distribution, with the delay spread (T_d) less than or equal to the guard interval (L), and that the inverse Fourier transform of the Doppler spectrum is the zeroth-order Bessel function of the first kind. The impulse response of the channel is described by $h(n,l)$, which denotes the tap gain of the lth tap at time n. In the simulations, each tap gain is independently generated by LPF of a white complex Gaussian process.

By the assumptions, the autocorrelation function of the channel is

$$E\left\{h(n_1,l_1)h^*(n_2,l_2)\right\} = c \cdot J_0\left(\frac{2\pi f_d T(n_1-n_2)}{N}\right) \cdot \mathrm{e}^{-l_1/L}\delta(l_1-l_2) \qquad (2)$$

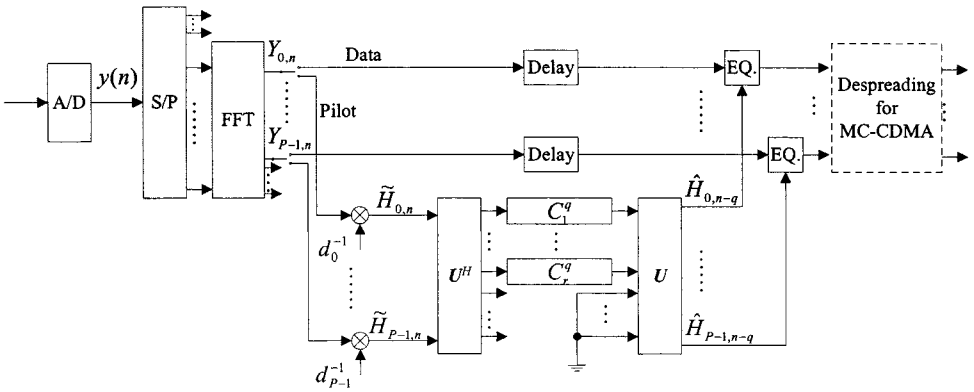


Fig. 2. Receiver structure: frequency-domain estimation and equalization.

where c, a normalization constant, is chosen to satisfy $c\sum_l \mathrm{e}^{-l/L}=1$, $J_0(\cdot)$ denotes the zeroth-order Bessel function of the first kind, and f_d is Doppler frequency in hertz. For the remainder of this paper, the number of taps is assumed to be $L+1$.

The received signal $y(n)$ can be expressed as

$$y(n)=\sum_{l=0}^{L} h(n,l)s(n-l)+w(n) \tag{3}$$

where $w(n)$ is AWGN. Hereafter, we assume $E_s=1$ and the variance of the AWGN is $\sigma^2=1/\mathrm{SNR}$, without loss of generality. Now, in the range $0\le n\le N-1$, the received signal is not corrupted by previous multicarrier symbols, due to the presence of the guard interval. Thus, in this interval the received signal becomes

$$\begin{aligned} y(n)=&\frac{1}{\sqrt{N}}\sum_{l=0}^{L} h(n,l)\sum_{k=0}^{P-1} d_k \exp\left[\frac{j2\pi(n-l)k}{N}\right]+w(n)\\ =&\frac{1}{\sqrt{N}}\sum_{k=0}^{P-1} d_k \mathrm{e}^{j2\pi nk/N}\sum_{l=0}^{L} h(n,l)\mathrm{e}^{-j2\pi lk/N}+w(n). \end{aligned} \tag{4}$$

By defining $H_k(n)\equiv\sum_{l=0}^{L} h(n,l)\mathrm{e}^{-j2\pi lk/N}$, which is the Fourier transform of the channel impulse response at time n, $y(n)$ can be written as

$$\begin{aligned} y(n)=&\frac{1}{\sqrt{N}}\sum_{k=0}^{P-1} d_k H_k(n)\mathrm{e}^{j2\pi nk/N}+w(n),\\ &0\le n\le N-1. \end{aligned} \tag{5}$$

Comparing this to the transmitted signal $s(n)$, we see that the time-varying multipath channel introduces a time-varying complex multiplier, $H_k(n)$, at each subcarrier.

C. Conventional Receiver Structure and ICI Analysis

Fig. 2 shows the conventional receiver structure including channel estimation. After excluding the ISI corrupted guard interval, the demodulation is performed using the FFT. In this figure, the second index, n, at the output of the FFT processor refers to the nth symbol processed, so that $Y_{k,n}$ is the output for the kth subcarrier at the nth symbol time. If we focus initially on the zeroth symbol and drop the second subscript, the kth subcarrier output from the FFT can be expressed as

$$Y_k=\frac{1}{\sqrt{N}}\sum_{n=0}^{N-1} y(n)\mathrm{e}^{-j2\pi nk/N}=d_kH_k+\alpha_k+W_k \tag{6}$$

where $H_k=1/N\sum_{n=0}^{N-1}H_k(n)$, $\alpha_k=1/N\sum_{m=0,m\neq k}^{P-1}d_m\sum_{n=0}^{N-1}H_m(n)\exp[j2\pi n(m-k)/N]$, and $W_k=1/\sqrt{N}\sum_{n=0}^{N-1}w(n)\mathrm{e}^{-j2\pi nk/N}$.

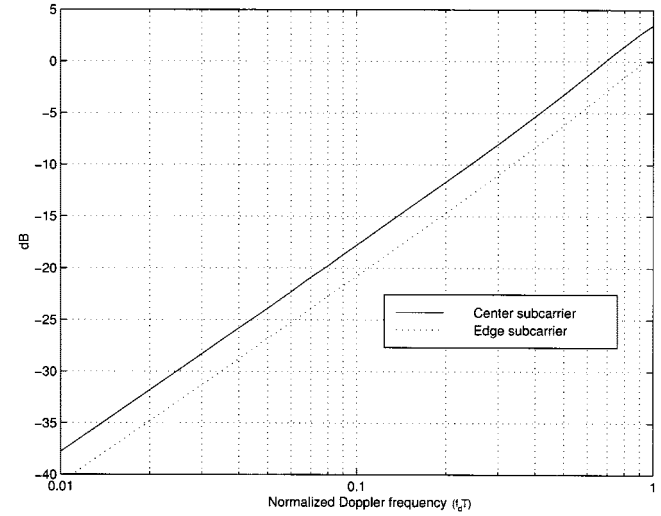
The α_k's represent interchannel interference (ICI) caused by the time-varying nature of the channel. In a time-invariant channel, one can see (by the orthogonality of the multicarrier basis waveforms) that α_k is zero, and $E\{|H_k|^2\}=1$. In a slowly time-varying channel (i.e., the normalized Doppler frequency f_dT is small), we can assume $E\{|H_k|^2\}\approx 1$, and $E\{|\alpha_k|^2\}\approx 0$. On the other hand, when the normalized Doppler frequency is high, the power of ICI cannot be ignored and, in addition, the power of the desired signal is reduced (i.e., $E\{|H_k|^2\}<1$). Therefore, we define the normalized ICI power as $E\{|\alpha_k|^2\}/E\{|H_k|^2\}$. In Appendix I the ICI power is evaluated, and it is shown that the ICI power is a function of the subcarrier index k. The center subcarrier experiences more ICI power than the edge subcarrier, as might be expected . Fig. 3(a) illustrates the normalized ICI power at both the center and edge subcarriers as a function of f_dT. The difference in the normalized ICI power is about 3 dB. There is a negligible ICI power difference with respect to the number of subcarriers, provided N is large enough, as shown in Fig. 3(b).

By assuming the normalized Doppler frequency is less than 0.02, the ICI power can be neglected compared to the background noise power. In this case, the signal at the output of the FFT processor for the nth OFDM symbol can be written as

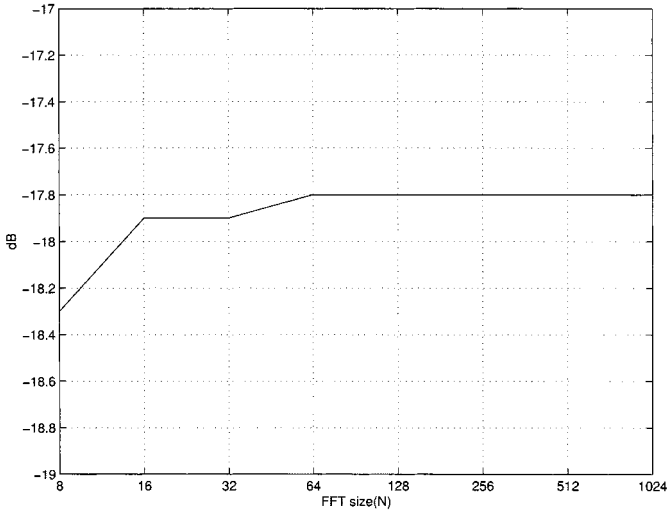
$$Y_{k,n}\approx d_{k,n}H_{k,n}+W_{k,n} \tag{7}$$

where k is the subcarrier index, $H_{k,n}=1/N\sum_{m=0}^{N-1}H_k(m+nN')$, and N' is $N+L$.

For channel estimation, insertion of pilot symbols is necessary. Several pilot patterns are possible. For ease of implementation, the two patterns shown in Fig. 4 may be considered. In Fig. 4(a), certain subcarriers are dedicated to pilot symbols, whereas in Fig. 4(b) all subcarriers in a given time slot are dedicated to pilot symbols. In the former scheme, the frequency-do-



(a)



(b)

Fig. 3. Normalized ICI power. (a) $N = 1024, P = 896$. (b) center subcarrier: $f_dT = 0.1, P = (7/8)N$.

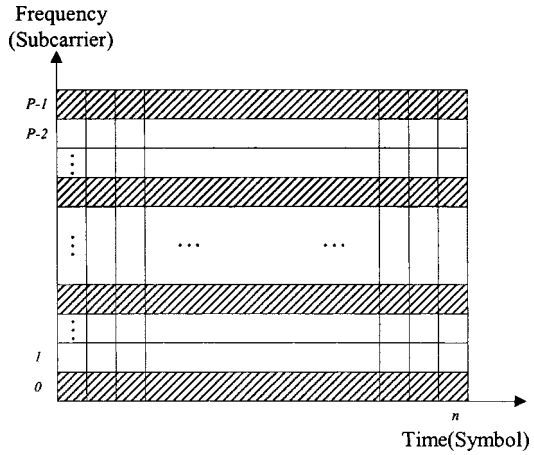
main correlation plays the key role in MMSE estimation, while the time-domain correlation is important in the latter. In a slowly time-varying channel, the time-domain correlation decays at a much slower rate than the frequency-domain correlation function. That means the pilot pattern of Fig. 4(b) requires a smaller number of pilot symbols than that of Fig. 4(a) for the same MSE. For the remainder of this paper, the pilot pattern of Fig. 4(b) is assumed.[1]

Using the pilot symbols, one can constitute a pilot channel vector for the nth multicarrier symbol as follows:

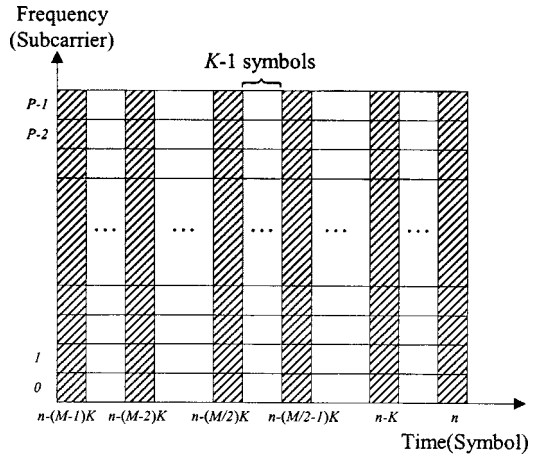
$$\widetilde{\boldsymbol{H}}_{:,n} = \begin{bmatrix} \widetilde{H}_{0,n} & \widetilde{H}_{1,n} & \dots & \widetilde{H}_{P-2,n} & \widetilde{H}_{P-1,n} \end{bmatrix}^{\mathrm{T}} \quad (8)$$

where the superscript $(\cdot)^{\mathrm{T}}$ denotes transpose and $\widetilde{H}_{k,n} = Y_{k,n}/d_{k,n} = H_{k,n} + W_{k,n}/d_{k,n}$. Assume that M pilot symbols are used for the estimation of the middle $K-1$ symbols in

[1]Although we assume symmetric pilot usage in this paper, one can use a time-asymmetric pattern in order to reduce the delay in the receiver.



(a)



(b)

Fig. 4. Pilot patterns. (a) Dedicated subcarriers. (b) Dedicated time slots.

Fig. 4(b). Define the column vector $\widetilde{\boldsymbol{H}}$ as the set of M pilot vectors

$$\widetilde{\boldsymbol{H}} = \begin{bmatrix} \widetilde{\boldsymbol{H}}^{\mathrm{T}}_{:,n} & \widetilde{\boldsymbol{H}}^{\mathrm{T}}_{:,n-K} & \dots & \widetilde{\boldsymbol{H}}^{\mathrm{T}}_{:,n-(M-2)K} & \widetilde{\boldsymbol{H}}^{\mathrm{T}}_{:,n-(M-1)K} \end{bmatrix}^{\mathrm{T}}. \quad (9)$$

By using a simple MMSE technique, we can obtain the best estimator in the MMSE sense as a linear combination of the pilot symbols as

$$\hat{\boldsymbol{H}}_{:,n-q} = C_q \widetilde{H} = R_{H_{:,n-q}\widetilde{H}} R^{-1}_{\widetilde{H}\widetilde{H}} \widetilde{H} \quad (10)$$

where $\boldsymbol{C}_q$ is a P-by-MP matrix and $((M/2)-1)K < q < M/2K$. The correlation matrices $\boldsymbol{R}_{H_{:,n-q}\widetilde{H}}$ and $\boldsymbol{R}_{\widetilde{H}\widetilde{H}}$ are defined as

$$\boldsymbol{R}_{H_{:,n-q}\widetilde{H}} = E\left\{ \boldsymbol{H}_{:,n-q} \widetilde{\boldsymbol{H}}^{\mathrm{H}} \right\} \quad (11)$$

and

$$\boldsymbol{R}_{\widetilde{H}\widetilde{H}} = E\left\{ \widetilde{\boldsymbol{H}} \widetilde{\boldsymbol{H}}^{\mathrm{H}} \right\} \quad (12)$$

where the superscript $(\cdot)^{\mathrm{H}}$ denotes conjugate transpose.

The correlation function of the $H_{k,n}$ for different time and frequency indices is

$$E\left\{H_{k+\Delta k,n+\Delta n}H_{k,n}^{*}\right\} = \frac{1}{N^2}\sum_{m_1=0}^{N-1}\sum_{m_2=0}^{N-1} J_0\left(\frac{2\pi f_d T(m_1-m_2+\Delta nN')}{N}\right) c \cdot \sum_{l} \mathrm{e}^{-l/L}\mathrm{e}^{-j2\pi\Delta kl/N}. \quad (13)$$

This function can be separated into the product of a time and frequency correlation as follows:

$$E\left\{H_{k+\Delta k,n+\Delta n}H_{k,n}^{*}\right\} = r_t(\Delta n)r_f(\Delta k) \quad (14)$$

where $r_t(\Delta n)=1/N^2\sum_{m_1=0}^{N-1}\sum_{m_2=0}^{N-1} J_0(2\pi f_d T(m_1-m_2+\Delta nN')/N)$, and $r_f(\Delta k) = c\sum_l \mathrm{e}^{-l/L}\mathrm{e}^{-j2\pi\Delta kl/N}$. This separation property makes possible a simple structure for the MMSE estimator as derived in Appendix II. In [7], the hard decision data as well as the pilots are used for channel estimation. Even with the error correction code, it inevitably suffers from the error propagation. Further, it requires much more taps of the estimator than using only the pilots [12].

III. Detection of Multicarrier Signals in a Fast Fading Channel

The conventional detection of multicarrier signals using the FFT exhibits relatively good performance at low values of f_dT. The performance of each subcarrier in Fig. 2 is very close to the theoretical bound of coherent detection of a single carrier in a flat fading channel. However, in an environment where the normalized Doppler frequency is high, there is an irreducible error floor even if all data are pilot symbols, since the pilot symbols themselves are corrupted by the ICI through the FFT. The FFT operation is simply a matched filter for each subcarrier. Since the time-varying channel destroys the orthogonality between subcarriers, the FFT introduces ICI. Therefore, the channel estimation and equalization should be performed immediately following the A/D conversion. In this section, we assume that the impulse response of the channel is known at each time and for each tap, and we discuss the detection problem assuming the channel is known. Channel estimation is addressed in Section IV.

The received signal $y(n)$ [see (5)] after excluding the guard interval can be expressed in a vector form as

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{d} + \boldsymbol{w} \quad (15)$$

where $\boldsymbol{y}=[y(0)\ \ y(1)\ \ldots\ y(N-1)]^{\mathrm{T}}$, $\boldsymbol{d}=[d_0\ \ d_1\ \ldots\ d_{P-1}]^{\mathrm{T}}$, $\boldsymbol{w}=[w(0)\ \ w(1)\ \ldots\ w(N-1)]^{\mathrm{T}}$, and the channel matrix $\boldsymbol{H}$ is given by (16), shown at the bottom of the page. Given the received vector $\boldsymbol{y}$ and the channel matrix $\boldsymbol{H}$, we can employ the following several detection methods.

A. Matched Filter (MF)

The received vector $\boldsymbol{y}$ can be decomposed in terms of the column vectors of $\boldsymbol{H}$ as follows:

$$\boldsymbol{y} = \boldsymbol{h}_0 d_0 + \boldsymbol{h}_1 d_1 + \cdots + \boldsymbol{h}_{P-1} d_{P-1} + \boldsymbol{w} \quad (17)$$

where $\boldsymbol{h}_k$ is the $(k+1)$th column vector of the matrix $\boldsymbol{H}$. In order to detect the data $\boldsymbol{d}_k$, the inner product is performed between the vector $\boldsymbol{y}$ and the vector $\boldsymbol{h}_k$. Therefore, the decision statistic $\boldsymbol{z}$ of the matched filter is given by

$$\boldsymbol{z} = \boldsymbol{H}^{\mathrm{H}}\boldsymbol{y} = \boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{d} + \boldsymbol{H}^{\mathrm{H}}\boldsymbol{w}. \quad (18)$$

Apparently, if the vectors $\boldsymbol{h}_k$ are mutually orthogonal, there is no interference since the P-by-P matrix $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}$ becomes an identity matrix. Unfortunately, this orthogonality does not hold in a time-varying channel. Hence, the matched filter suffers from ICI.

B. Least Square

The linear model (15) leads to the classical least squares problem [13]. The least squares detection statistic is given by

$$\boldsymbol{z} = \boldsymbol{H}^{+}\boldsymbol{y} \quad (19)$$

where the P-by-N matrix $\boldsymbol{H}^{+} = (\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{H}}$ is the pseudoinverse of $\boldsymbol{H}$.

C. Minimum Mean Squared Error (MMSE)

The MMSE detector chooses the equalizer matrix $\boldsymbol{G}^{\mathrm{H}}$ which minimizes the cost function $E\left\{|\boldsymbol{d}-z|^2\right\}$ where $z=\boldsymbol{G}^{\mathrm{H}}\boldsymbol{y}$. The resulting detection statistic becomes

$$z = \boldsymbol{G}^{\mathrm{H}}\boldsymbol{y} = \boldsymbol{H}^{\mathrm{H}}(\boldsymbol{H}H^{\mathrm{H}} + \sigma^2\boldsymbol{I}_N)^{-1}\boldsymbol{y} \quad (20)$$

where $\boldsymbol{I}_N$ is the N-by-N identity matrix. As opposed to least squares, the MMSE detector requires the knowledge of the noise power. The equalizer matrix $\boldsymbol{G}^{\mathrm{H}}$ satisfies the following identity, proved in Appendix III:

$$\boldsymbol{G}^{\mathrm{H}} = \boldsymbol{H}^{\mathrm{H}}\left(\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}} + \sigma^2\boldsymbol{I}_N\right)^{-1} = \left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H} + \sigma^2\boldsymbol{I}_P\right)^{-1}\boldsymbol{H}^{\mathrm{H}}. \quad (21)$$

The right-hand side of (21) is very similar to the pseudoinverse of $\boldsymbol{H}$ in the least squares except the identity matrix accounts for the noise power. The insertion of the noise power in the inverse

$$\boldsymbol{H} = \frac{1}{\sqrt{N}}\begin{bmatrix} H_0(0) & H_1(0) & \cdots & H_{P-1}(0) \\ H_0(1) & H_1(1)\mathrm{e}^{j2\pi/N} & \cdots & H_{P-1}(1)\mathrm{e}^{j2\pi(P-1)/N} \\ \vdots & \vdots & \cdots & \vdots \\ H_0(N-1) & H_1(N-1)\mathrm{e}^{j2\pi(N-1)/N} & \cdots & H_{P-1}(N-1)\mathrm{e}^{j2\pi(P-1)(N-1)/N} \end{bmatrix} \quad (16)$$

matrix of (21) reduces the noise enhancement. From (A28) we can express the least squares in expanded form

$$\boldsymbol{G}^{\mathrm{H}} = \boldsymbol{H}^{+} = \sum_{i=1}^{r} \frac{1}{\lambda_i} \boldsymbol{v}_i \boldsymbol{u}_i^{\mathrm{H}}. \tag{22}$$

When the singular value λ is small, one can expect that the least squares experiences more noise enhancement than the MMSE detection.

D. MMSE with Successive Detection (SD)

The decision statistic z has two terms in general

$$z = \boldsymbol{G}^{\mathrm{H}} \boldsymbol{H} \boldsymbol{d} + \boldsymbol{G}^{\mathrm{H}} \boldsymbol{w}. \tag{23}$$

The first term is the data component and the other is the noise component. In the LS technique, the product $\boldsymbol{G}^{\mathrm{H}}\boldsymbol{H}$ is an identity matrix. This means there is no ICI after the equalization. However, as mentioned earlier, this technique suffers from noise enhancement. The noise enhancement increases as the normalized Doppler frequency gets higher because the minimum nonzero singular value, $\lambda_{\min}$, becomes smaller. On the other hand, even if the MMSE detection technique generates some residual interference, it provides a better balance in minimizing the sum of the noise enhancement and the residual interference.

Now, as the Doppler frequency increases, the channel becomes time-selective. From another point of view, it provides time diversity, since the data d_k is carried on N random variables (i.e., the N elements of $\boldsymbol{h}_k$ in one symbol duration) whose correlation can be small. That is, the rapidly time-varying channel not only destroys the orthognality, but also provides us with time diversity. Certainly, if there is only one subcarrier, all previous methods can take advantage of the time diversity. However, in multicarrier systems, the MF generates ICI, while the LS causes noise enhancement. The MMSE detection, on the other hand, should be able to make good use of the time selectivity since it minimizes both the residual interference and the noise enhancement while retaining the gain from the time diversity. Thus, one might expect that the performance would be better as the normalized Doppler frequency becomes large. However, the residual interference and the noise enhancement grow as well. Therefore, the gain from the time-selective channel can be overwhelmed by the residual interference and the noise enhancement, for some normalized Doppler frequencies. This effect is more severe for higher order modulations. That is, the degradation from the interference and the noise enhancement is more dominant than the advantage from the time-selective channel for higher order modulations.

In order to fully utilize the time diversity while suppressing the residual interference and the noise enhancement, we detect the data one-by-one instead of detecting all the data simultaneously, as in the previous methods. Hence, we adopt the successive detection technique which is widely used in DS-CDMA systems for the multi-user detection and has also been implemented in [14] in order to achieve higher spectral efficiency.

Consider first the SD technique from [14]. We first detect the data, d_k, which has the highest post-detection SNR among undetected data, assuming LS detection. Equivalently choose the $(k+1)$th column vector of the equalizer matrix $\boldsymbol{G}$ which satisfies the condition

$$\begin{aligned} \arg\max_k \mathrm{SNR}_k &= \arg\max_k \frac{|\langle \boldsymbol{g}_k, \boldsymbol{h}_k \rangle|^2}{\sigma^2 ||\boldsymbol{g}_k||^2} \\ &= \arg\max_k \frac{1}{\sigma^2 ||\boldsymbol{g}_k||^2} \leftrightarrow \arg\min_k ||\boldsymbol{g}_k||^2 \end{aligned} \tag{24}$$

where $\boldsymbol{g}_k$ is the $(k+1)$th column vector of the equalizer matrix $\boldsymbol{G}$ (the inner product, $\boldsymbol{g}_k^{\mathrm{H}}\boldsymbol{h}_k$, is unity, because of the orthogonality condition in the assumed LS detection). After making a hard decision, the received vector $\boldsymbol{y}$ is modified to

$$\boldsymbol{y}_{\mathrm{new}} = \boldsymbol{y}_{\mathrm{old}} - \boldsymbol{h}_k \hat{d}_k \tag{25}$$

where $\hat{d}_k$ is the hard decision data. The column vector $\boldsymbol{h}_k$ of the channel matrix $\boldsymbol{H}$ is replaced by a zero vector, and then, the corresponding equalizer matrix $\boldsymbol{G}^{\mathrm{H}}$ is updated. This successive detection is analogous to decision feedback equalization in that one can expect an error propagation phenomenon. As long as this hard decision data is correct, the new vector $\boldsymbol{y}_{\mathrm{new}}$ has fewer interferers. With the smaller number of interferers, higher post-detection SNR can be achieved [14]. The ordering influences the overall performance, and the ordering is chosen by the post-detection SNR as described above. This choice is optimal when the multiplication $\boldsymbol{G}^{\mathrm{H}}\boldsymbol{H}$ becomes an identity matrix as in the LS. However, there is residual interference through the equalizer in the MMSE case. In this case, therefore, we determine the detection order by the post-detection signal-to-interference and noise power ratio (SINR) based upon MMSE detection. For a particular data symbol, d_k, the SINR is defined as follows:

$$\mathrm{SINR}_k = \frac{|\langle \boldsymbol{g}_k, \boldsymbol{h}_k \rangle|^2}{\sum_{m, m \neq k} |\langle \boldsymbol{g}_k, \boldsymbol{h}_m \rangle|^2 + \sigma^2 ||\boldsymbol{g}_k||^2}. \tag{26}$$

The detailed detection procedure is described in Fig. 5.

E. Performance Comparisons

By assuming each interference is uncorrelated Gaussian r.v., we can calculate the theoretical symbol error rate (SER) of the above detection techniques. The probability of symbol error of BPSK is given by

$$\begin{aligned} P[e] = &\frac{1}{P} \sum_k \int \cdots \int p\left(h(0,0), \ldots, h((N-1), L)\right) \\ &\cdot Q\left(\sqrt{2\mathrm{SINR}_k}\right) dh(0,0) \cdots dh((N-1), L). \end{aligned} \tag{27}$$

Instead of finding a closed form of the SER, we evaluate the SER by a Monte Carlo integration. The channel impulse response $h(n,l)$ is generated through Doppler spectrum generator. The corresponding SINR_k at each subcarrier is calculated, substituted into the $Q(\cdot)$ function, and the sum of the trials is averaged accordingly. In the SD, the lower bound can be obtained by the assumption of ideal decision feedback.

$j = 1$

$\boldsymbol{G}^H = (\boldsymbol{H}^H\boldsymbol{H} + \sigma^2\boldsymbol{I}_P)^{-1}\boldsymbol{H}^H$

$$i_1 = \arg\max_k \ SINR_k = \arg\max_k \ SINR_k = \frac{\left|\langle \boldsymbol{g}_k, \boldsymbol{h}_k \rangle\right|^2}{\sum_{m,m\neq k}\left|\langle \boldsymbol{g}_k, \boldsymbol{h}_m \rangle\right|^2 + \sigma^2\|\boldsymbol{g}_k\|^2}$$

Loop

$z_{i_j} = \boldsymbol{g}_{i_j}^H \boldsymbol{y}$

$\hat{d}_{i_j} = hard\ decision\ of\ z_{i_j}$

$\boldsymbol{y} = \boldsymbol{y} - \boldsymbol{h}_{i_j}\hat{d}_{i_j}$

$\boldsymbol{H} = \left[\boldsymbol{h}_0 \ \cdots \ \boldsymbol{h}_{i_j-1} \ \boldsymbol{0} \ \boldsymbol{h}_{i_j+1} \ \cdots \ \boldsymbol{h}_{P-1}\right]$

$\boldsymbol{G}^H = (\boldsymbol{H}^H\boldsymbol{H} + \sigma^2\boldsymbol{I}_P)^{-1}\boldsymbol{H}^H$

$$i_{j+1} = \arg\max_{k\in\{i_1,\cdots,i_j\}} \frac{\left|\langle \boldsymbol{g}_k, \boldsymbol{h}_k \rangle\right|^2}{\sum_{m,m\in\{i_1,\cdots,i_j\},m\neq k}\left|\langle \boldsymbol{g}_k, \boldsymbol{h}_m \rangle\right|^2 + \sigma^2\|\boldsymbol{g}_k\|^2}$$

$j = j + 1$

Fig. 5. MMSE with the SD.

In the following, we demonstrate, through extensive computer simulations, the relative performance of the above mentioned detection techniques for OFDM systems. For simulation purposes only, we have chosen the number of subcarriers (FFT size) as 32. The number of active subcarriers is 28 and the normalized delay spread T_d/T is 1/16. In this particular case, the number of paths becomes 3. The delay power spectrum has an exponential distribution. In Fig. 6(a) and (b), the SERs of BPSK and 16PSK, respectively, are shown as a function of SNR when the normalized Doppler frequency f_dT is 0.1. The MF suffers from severe ICI. The LS detection has a smaller error floor than the MF. In BPSK, the SER performance of the MMSE is better than the theoretical bound of BPSK in a single carrier with ideal channel estimation operating in a flat fading channel, while in 16PSK the performance is slightly worse than the coherent 16PSK. The MMSE with the SD has the best performance in both cases. Due to the time diversity, the SER performance of the MMSE with SD starts to improve for SNR above 6 dB compared to the theoretical bound of coherent BPSK, while in 16PSK the performance of the MMSE with the SD gets better for SNR exceeding 22 dB due to an error propagation at low SNR. When f_dT is 1.0 in Fig. 6(c), the improvement of the SD is apparent whereas all other techniques experience ICI or noise enhancement. Fig. 6(d) illustrates the SER performances of 16PSK at 30 dB SNR as a function of the Doppler frequency. Only the MMSE detection with SD can exploit the time-varying channel at the higher order modulations while the others increase the error floor. To confirm this characteristic, the capacity of the time-varying multipath channel and the information rate of MMSE detection with SD are explored in [15]. Note the error propagation of the MMSE with SD is more apparent at higher f_dT.

IV. Channel Estimation

In the previous section, we have investigated several detection methods assuming the ideal channel response is available. It seems the estimation of the N-by-P channel matrix $\boldsymbol{H}$ is impossible even with pilot symbols since there are more unknowns to be determined than known equations. Thus, there is no way to recover the NP unknown elements of $\boldsymbol{H}$ even though the pilot data vector $\boldsymbol{d}$ is known. In what follows, the estimation of the impulse response of the channel is described.

A. Channel Estimation

Suppose that the channel is stationary and its statistics (at least the worst conditions) are known. Basically, the channel estimation is an MMSE technique relying on pilot symbols. However, we do not make any attempt to estimate the channel matrix $\boldsymbol{H}$ at each pilot position and interpolate the matrix to get a channel matrix at a particular time instant. Instead, we directly estimate the impulse response $h(n,l)$ as a linear combination of the received signal $y(n)$. The received pilot signals to be used by the estimator are now defined as a vector

$$\boldsymbol{y}_P = \left[\boldsymbol{y}_n^{\mathrm{T}} \quad \boldsymbol{y}_{n-K}^{\mathrm{T}} \quad \cdots \quad \boldsymbol{y}_{n-(M-2)K}^{\mathrm{T}} \quad \boldsymbol{y}_{n-(M-1)K}^{\mathrm{T}}\right]^{\mathrm{T}} \tag{28}$$

where $\boldsymbol{y}_n = [y(nN') \ \ y(1+nN') \ \ \ldots \ \ y(N-1+nN')]^{\mathrm{T}}$ is the nth symbol. Define the channel vector to be estimated as (29), shown at the bottom of the page, where $\boldsymbol{h}_n = [h(nN',0) \ \ldots \ h(N-1+nN',0) \ \ldots \ h(nN',L) \ \ldots \ h(N-1+nN',L)]^{\mathrm{T}}$ is now the collection of the impulse response of each tap at the nth symbol (note that the notation $\boldsymbol{h}_n$ has a different meaning than in Section III). The MMSE estimator leads to the following linear equation:

$$\hat{h} = C \cdot y_P = \boldsymbol{R}_{hy_P}\boldsymbol{R}_{y_Py_P}^{-1}\boldsymbol{y}_P. \tag{30}$$

The autocorrelation matrix $\boldsymbol{R}_{y_Py_P}$ and the cross-correlation matrix $\boldsymbol{R}_{hy_P}$ can be obtained from (2) and (3) since the transmitted pilot signal $s(n)$ is known. In the time-domain MMSE channel estimator, the MSE performance is highly dependent on the pilot signal waveform. The pilot signal design in the time-varying Rayleigh fading channel is another challenging topic, which is beyond the scope of this paper. Hereafter, we use the usual OFDM signal as a pilot signal $s(n)$, but the pilot data d_k are chosen by making use of the average normalized MSE (NMSE) performance of the estimator. The NMSE is defined as

$$\mathrm{NMSE}(n) = \frac{\sum_l E\left\{\left|h(n,l) - \hat{h}(n,l)\right|^2\right\}}{\sum_l E\left\{|h(n,l)|^2\right\}}. \tag{31}$$

$$\boldsymbol{h} = [\boldsymbol{h}_{n-(M/2-1)K+1}^{\mathrm{T}} \quad \boldsymbol{h}_{n-(M/2-1)K+2}^{\mathrm{T}} \quad \cdots \quad \boldsymbol{h}_{n-(M/2-1)K+K-2}^{\mathrm{T}} \quad \boldsymbol{h}_{n-(M/2-1)K+K-1}^{\mathrm{T}}]^{\mathrm{T}} \tag{29}$$

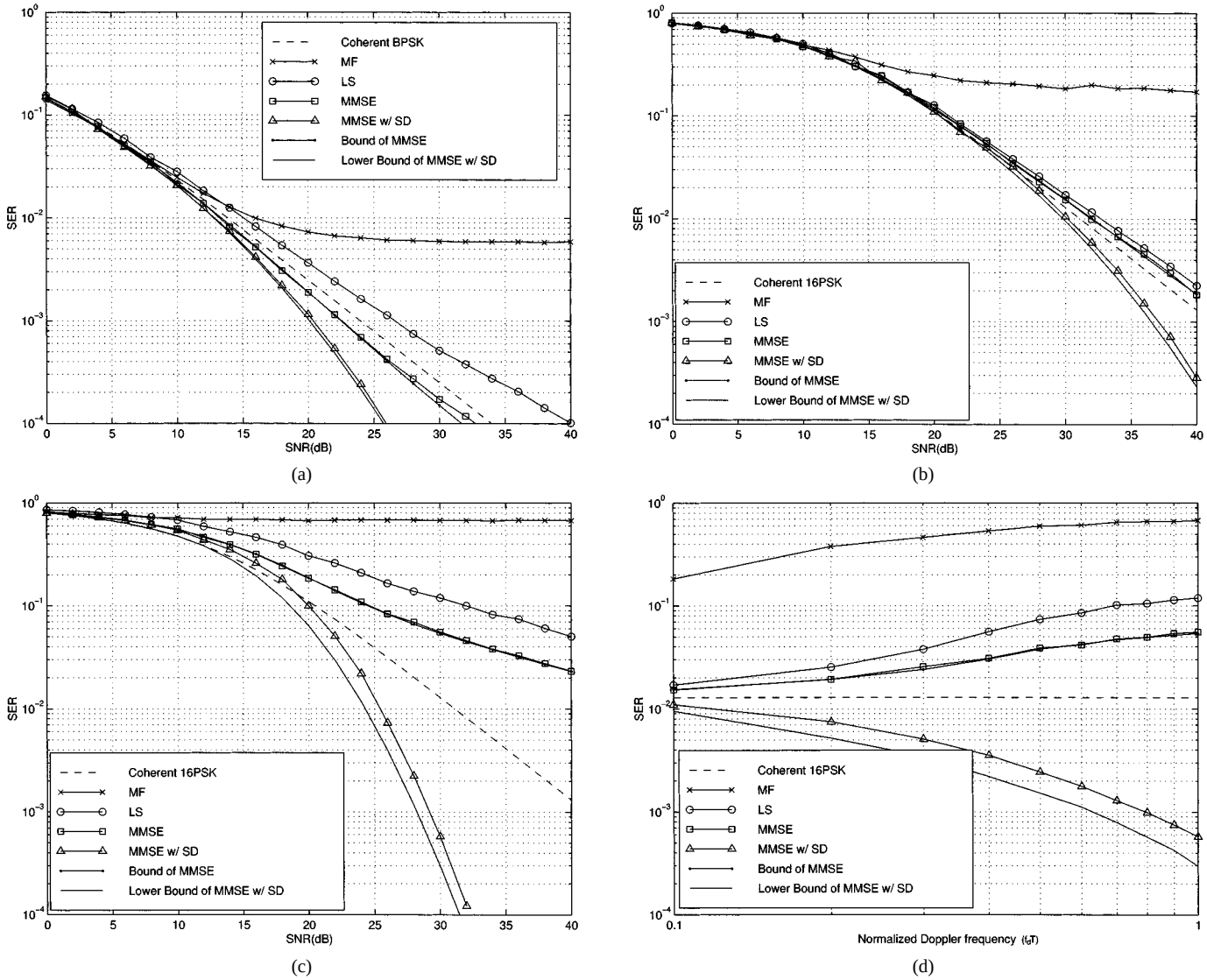


Fig. 6. Performance comparisons with ideal channel estimation: $T_d/T = 1/16$, $N = 32$, $P = 28$. (a) BPSK: $f_dT = 0.1$. (b) 16PSK: $f_dT = 0.1$. (c) 16PSK: $f_dT = 1.0$. (d) 16PSK: $\mathrm{SNR} = 30$ dB.

Certainly, this is a function of the time over $(K-1)$ symbols. The average NMSE $(\overline{\mathrm{NMSE}})$ is given by

$$\overline{\mathrm{NMSE}} = \sum_{q=0}^{K-2} \sum_{m=0}^{N-1} \frac{\mathrm{NMSE}(qN'+m)}{(K-1)N}. \tag{32}$$

The behavior of the theoretical $\overline{\mathrm{NMSE}}$ with different pilot spacing (K) and number of pilot symbols (M) is shown in Fig. 7 when the normalized Doppler frequency f_dT is 0.1. It is natural that the larger pilot spacing requires a larger number of pilot symbols. The simulation results for the SER are shown in Fig. 8(a) with $f_dT = 0.1$, $K = 3$ and $M = 6$. The conventional frequency-domain method has an error floor. There is no difference in the SER performance between the ideal channel estimation and pilot-assisted estimation in each detection method up to 22 dB in SNR. When the SNR is below 22 dB, the $\overline{\mathrm{NMSE}}$ is negligible since the $\overline{\mathrm{NMSE}}$ is 10 dB lower than the noise power. Fig. 8(b) shows the theoretical $\overline{\mathrm{NMSE}}$ together with the simulation result. The deviation of the simulation result from the theoretical bound at high SNR comes from the nonideal Doppler spectrum generation. Fig. 9 depicts the SER and the $\overline{\mathrm{NMSE}}$ performances when $f_dT = 0.3$, $K = 2$ and $M = 12$.

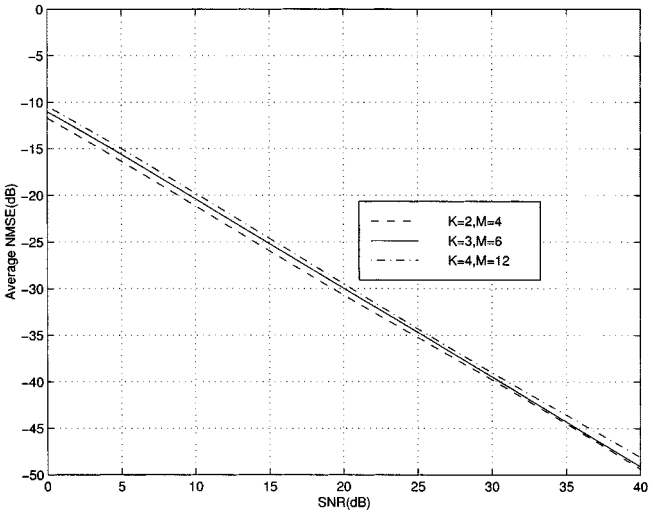
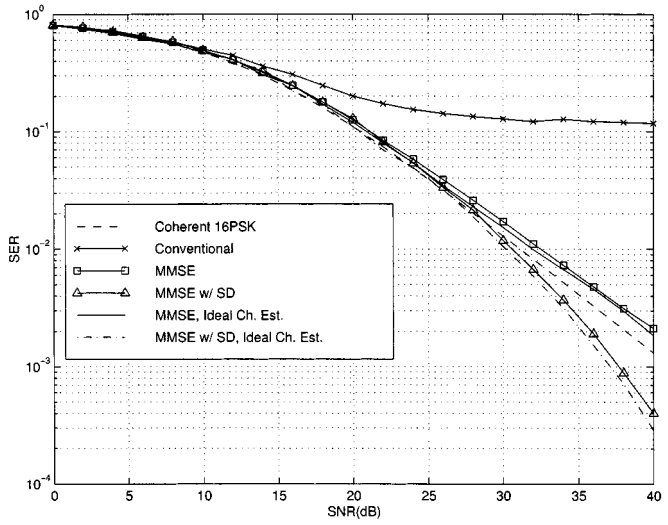
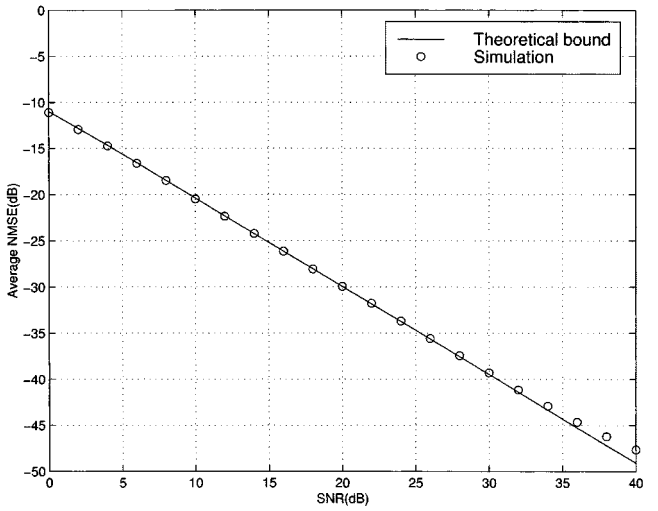


Fig. 7. Theoretical bounds of $\overline{\mathrm{NMSE}}$ with different pilot spacings (K) and number of pilots (M): $f_dT = 0.1$, $T_d/T = 1/16$, $N = 32$, $P = 28$.

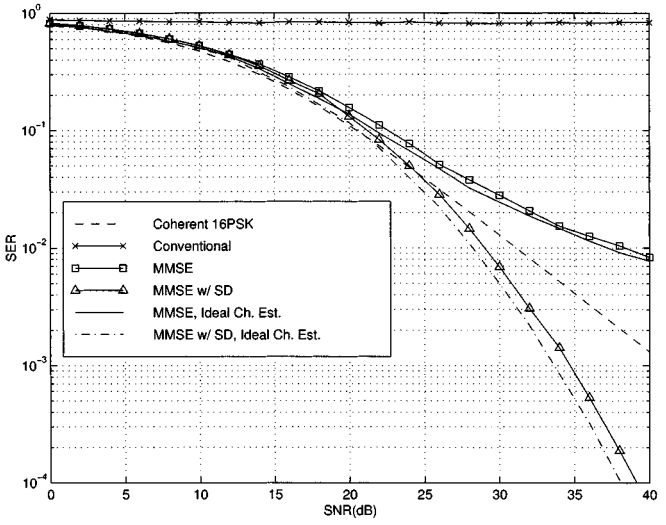


(a)

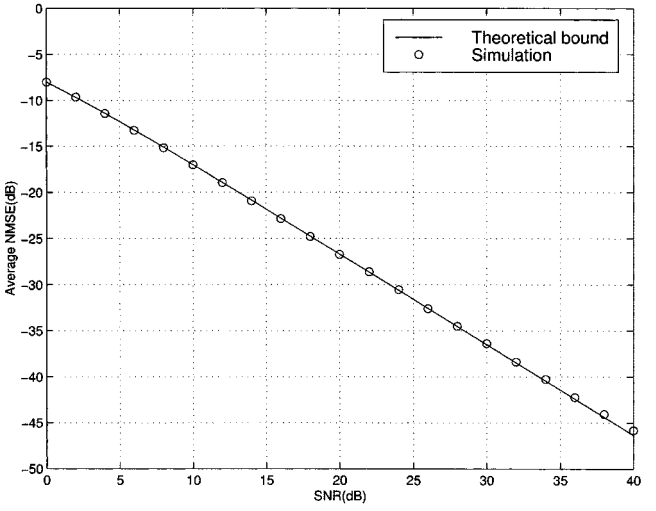


(b)

Fig. 8. Simulations with pilot assisted channel estimation: $K = 3$, $M = 6$, $f_dT = 0.1$, $T_d/T = 1/16$, $N = 32$, $P = 28$. (a) SER. (b) $\overline{\text{NMSE}}$.



(a)



(b)

Fig. 9. Simulations with pilot assisted channel estimation: $K = 2$, $M = 12$, $f_dT = 0.3$, $T_d/T = 1/16$, $N = 32$, $P = 28$. (a) SER. (b) $\overline{\text{NMSE}}$.

B. Low Rank Approximation

The size of the channel estimator C in (30) is $(K-1)(L+1)N$-by-NM. The required number of complex multiplications is $(K-1)(L+1)MN^2$ for estimating $K-1$ symbols. A reduced complexity MMSE estimator can be achieved by employing optimal rank reduction [13]. First, let us write $\boldsymbol{C}\boldsymbol{R}_{y_Py_P}^{1/2}$ in a singular value decomposition (SVD) representation:

$$\boldsymbol{C}\boldsymbol{R}_{y_Py_P}^{1/2} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{H}} \tag{33}$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are unitary matrices, and the diagonal matrix $\boldsymbol{\Lambda}$ is given by

$$\boldsymbol{\Lambda} = \begin{bmatrix} \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_D) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}. \tag{34}$$

After replacing the last $D-r$ singular values by zero, the low rank estimator $\boldsymbol{C}_r$ becomes

$$\boldsymbol{C}_r = \boldsymbol{U} \begin{bmatrix} \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \widetilde{\boldsymbol{V}}^{\mathrm{H}} \tag{35}$$

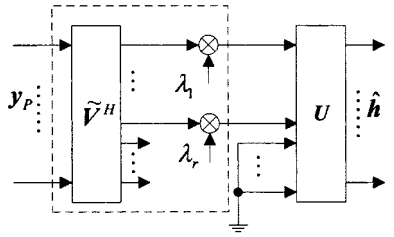
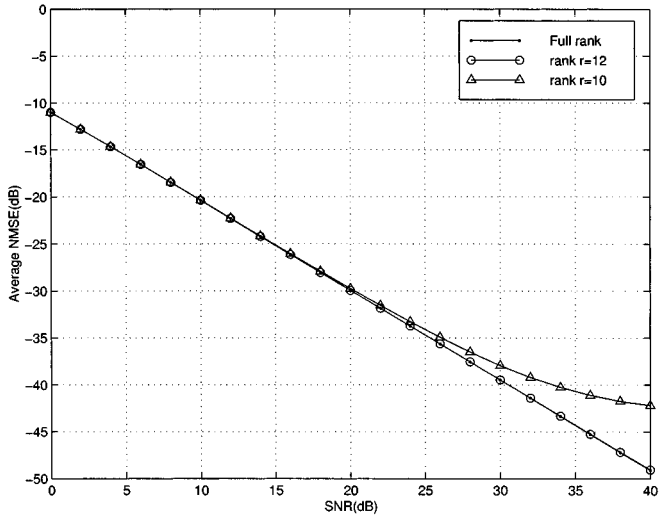
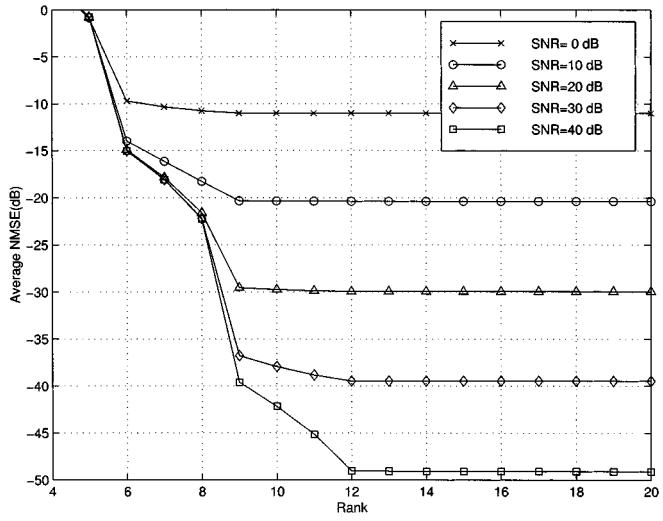


Fig. 10. Low rank estimator.

where $\widetilde{\boldsymbol{V}}^{\mathrm{H}} = \boldsymbol{V}^{\mathrm{H}}\boldsymbol{R}_{y_Py_P}^{-1/2}$. The additional MSE caused by the low rank approximation is incorporated in the sum $\sum_{i=r+1}^{D}\lambda_i$. The implementation of the low rank estimator is illustrated in Fig. 10. Note the singular value λ can be included in either $\boldsymbol{U}$ or $\widetilde{\boldsymbol{V}}$. Then, the number of complex multiplications becomes $r(K-1)(L+1)N + rNM$. Fig. 11(a) shows the $\overline{\text{NMSE}}$ when r is equal to 10 and 12. The $\overline{\text{NMSE}}$ is plotted as a function of the rank for several SNR values in Fig. 11(b). In this particular case, there is no difference in $\overline{\text{NMSE}}$ between the full rank



(a)



(b)

Fig. 11. Theoretical bounds of $\overline{\mathrm{NMSE}}$ as a function of the rank: $f_dT = 0.1$, $T_d/T = 1/16$, $N = 32$, $P = 28$, $K = 3$, $M = 6$.

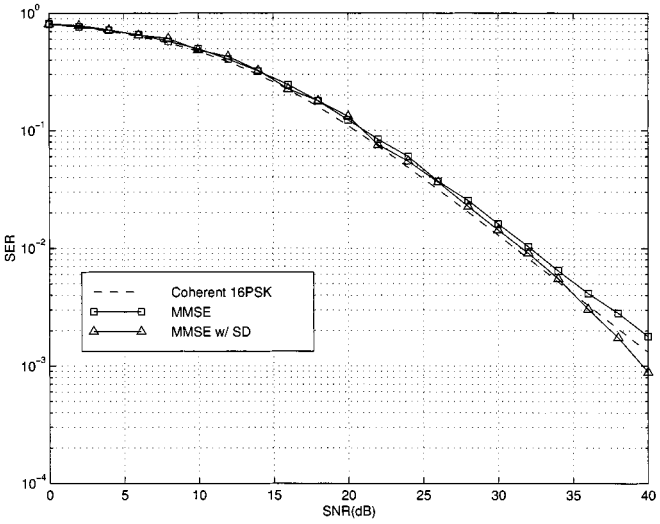
and the low rank $(r = 12)$ estimators. When the rank r is 12, the number of multiplications is reduced to 4,608 from 36 864 without losing $\overline{\mathrm{NMSE}}$ and SER performance.

C. Mismatch

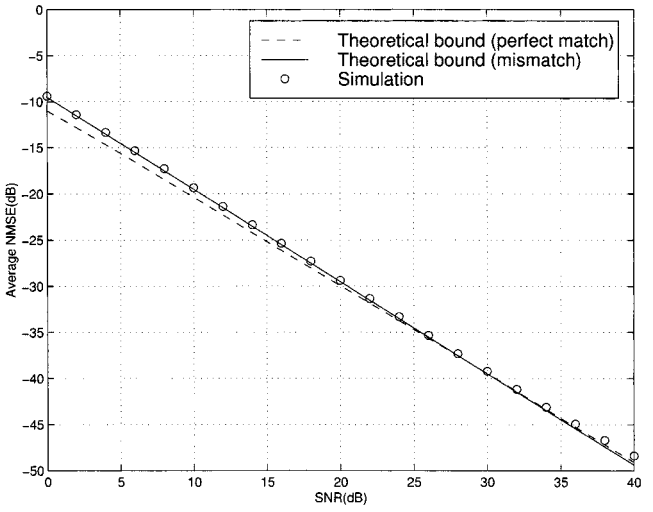
The MMSE channel estimator described above requires knowledge of the channel statistics, i.e., the noise power, Doppler spread, delay spread and multipath intensity profile. In practice, the true channel characteristic is almost always unknown. Estimating channel parameters is another significant problem, and accurate estimation might not be possible. Instead, we investigate the $\overline{\mathrm{NMSE}}$ of the estimator, which is designed for the worst-case condition, when there is a mismatch between assumed channel parameters and true channel characteristics. As described in Table I, the assumed channel differs significantly from the actual channel. Fig. 12 illustrates the mismatch simulation results of the low rank estimator. Since the actual Doppler frequency is smaller than that of the assumed channel, there is a smaller gain in the MMSE with the SD.

TABLE I
MISMATCH CONDITIONS

	Assumed channel conditions	Actual channel conditions
SNR	40 dB	0~40 dB
f_dT	0.1	0.05
Delay spread (T_d / T)	1/16	1/32
Multipath intensity profile	Exponential distribution	Uniform distribution



(a)



(b)

Fig. 12. Simulations of mismatch in Rayleigh fading channel. Low rank $r = 12$, $K = 3$, $M = 6$, $N = 32$, $P = 28$. (a) SER. (b) $\overline{\mathrm{NMSE}}$.

V. CONCLUSION

Operation over a time-varying channel results in an error floor when using frequency-domain estimation and equalization in multicarrier systems. We have presented several time-domain techniques for detection and estimation. The MMSE detection technique is able to exploit the time-varying channel as a source

of time diversity. However, it still results in residual interference, causing performance degradation for higher order modulations. In order to effectively suppress both the residual interference and the noise enhancement, MMSE with successive detection is used. The MMSE technique with the SD permits us to achieve time diversity even for higher order modulations. If this detection technique is adopted in MC-CDMA, we can make full use of the time-varying and frequency-selective channel as a source of time and frequency diversity [15].

With regard to channel estimation, we have investigated the performance of the time-domain MMSE estimator. The low rank estimator is shown to be a robust estimator to changes in the channel characteristics.

APPENDIX I

Assuming the data on each subcarrier is uncorrelated, and $E\left\{|d_m|^2\right\}=1$, the ICI power becomes

$$\begin{aligned}E\left\{|\alpha_k|^2\right\}&=\frac{1}{N^2}\sum_{m=0,m\neq k}^{P-1}E\left\{|d_m|^2\right\}\sum_{n_1=0}^{N-1}\sum_{n_2=0}^{N-1}E\left\{H_m(n_1)H_m^*(n_2)\right\}\\&\quad\cdot\exp\left[\frac{j2\pi(n_1-n_2)(m-k)}{N}\right]\\&=\frac{1}{N^2}\sum_{m=0,m\neq k}^{P-1}\sum_{n_1=0}^{N-1}\sum_{n_2=0}^{N-1}E\left\{H_m(n_1)H_m^*(n_2)\right\}\\&\quad\cdot\exp\left[\frac{j2\pi(n_1-n_2)(m-k)}{N}\right].\end{aligned}\tag{A1}$$

The autocorrelation of the frequency response $\left(H_m(n)\equiv\sum_{l=0}^{L}h(n,l)e^{-j2\pi lm/N}\right)$ of the channel is

$$\begin{aligned}E\left\{H_m(n_1)H_m^*(n_2)\right\}&=\sum_{l_1=0}^{L}\sum_{l_2=0}^{L}E\left\{h(n_1,l_1)h^*(n_2,l_2)\right\}\\&\quad\cdot\exp\left[\frac{-j2\pi m(l_1-l_2)}{N}\right].\end{aligned}\tag{A2}$$

From the (2), it becomes

$$\begin{aligned}E\left\{H_m(n_1)H_m^*(n_2)\right\}&=\sum_{l=0}^{L}E\left\{h(n_1,l)h^*(n_2,l)\right\}\\&=c\sum_{l=0}^{L}J_0\left(\frac{2\pi f_dT(n_1-n_2)}{N}\right)e^{-l/L}\\&=J_0\left(\frac{2\pi f_dT(n_1-n_2)}{N}\right).\end{aligned}\tag{A3}$$

As shown above, it is independent of the delay power spectrum. Substituting (A3) into (A1), the ICI power can be written as

$$\begin{aligned}E\left\{|\alpha_k|^2\right\}=&\frac{1}{N^2}\sum_{m=0,m\neq k}^{P-1}\sum_{n_1=0}^{N-1}\sum_{n_2=0}^{N-1}J_0\left(\frac{2\pi f_dT(n_1-n_2)}{N}\right)\\&\cdot\exp\left[\frac{j2\pi(n_1-n_2)(m-k)}{N}\right].\end{aligned}\tag{A4}$$

The fact that $J_0(\cdot)$ is an even function simplifies the above calculation as

$$\begin{aligned}&E\left\{|\alpha_k|^2\right\}\\&=\frac{1}{N^2}\sum_{m=0,m\neq k}^{P-1}\left(N+2\sum_{n=1}^{N-1}(N-n)J_0\right.\\&\qquad\left.\cdot\left(\frac{2\pi f_dTn}{N}\right)\cos\left(\frac{2\pi n(m-k)}{N}\right)\right).\end{aligned}\tag{A5}$$

With the same way, we can obtain

$$E\left\{|H_k|^2\right\}=\frac{1}{N^2}\left(N+2\sum_{n=1}^{N-1}(N-n)J_0\left(\frac{2\pi f_dTn}{N}\right)\right).\tag{A6}$$

Therefore, the normalized ICI power is (A7), shown at the bottom of the page.

APPENDIX II

The cross-correlation matrix between the two pilot vectors at the time n_1 and n_2 is

$$E\left\{\widetilde{\boldsymbol{H}}_{:,n_1}\widetilde{\boldsymbol{H}}_{:,n_2}^{\mathrm{H}}\right\}=r_t(n_1-n_2)\boldsymbol{R}_f+\sigma^2\delta(n_1-n_2)\boldsymbol{I}_P\tag{A8}$$

where the scalar valued time-domain correlation function $r_t(\cdot)$ is defined in (14), the P-by-P frequency domain matrix $\boldsymbol{R}_f$ is defined as

$$R_f=\begin{bmatrix}r_f(0)&r_f(-1)&\cdots&r_f(-P+1)\\r_f(1)&r_f(0)&\cdots&r_f(-P+2)\\\vdots&\vdots&\ddots&\vdots\\r_f(P-1)&r_f(P-2)&\cdots&r_f(0)\end{bmatrix}\tag{A9}$$

with $\boldsymbol{I}_P$ is a P-by-P identity matrix, and the power of the pilot data is assumed to be equal to the power of the information data, otherwise the noise power should be scaled accordingly. Hence, the autocorrelation matrix $\boldsymbol{R}_{\widetilde{H}\widetilde{H}}$ (12) can be expressed as (A10), shown at the bottom of the next page. Note $r_t(k)=r_t(-k)$ and the frequency-domain matrix $\boldsymbol{R}_f$ is a Hermitian matrix since $r_f(k)=r_f^*(-k)$. Then, we have the following representation of $\boldsymbol{R}_f$ by the SVD

$$\boldsymbol{R}_f=\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}}\tag{A11}$$

$$\frac{E\left\{|\alpha_k|^2\right\}}{E\left\{|H_k|^2\right\}}=\frac{\sum_{m=0,m\neq k}^{P-1}\left(N+2\sum_{n=1}^{N-1}(N-n)J_0\left(\frac{2\pi f_dTn}{N}\right)\cos\left(\frac{2\pi n(m-k)}{N}\right)\right)}{\left(N+2\sum_{n=1}^{N-1}(N-n)J_0\left(\frac{2\pi f_dTn}{N}\right)\right)}\tag{A7}$$

where $\boldsymbol{U}$ is a unitary matrix containing the singular vectors and $\boldsymbol{\Lambda}$ is a diagonal matrix whose elements are the singular values $\lambda_1 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_P = 0$. Then the inverse matrix $\boldsymbol{R}^{-1}_{\widetilde{H}\widetilde{H}}$ becomes

$$\boldsymbol{R}^{-1}_{\widetilde{H}\widetilde{H}} = \begin{bmatrix} \boldsymbol{U} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{U} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{U} \end{bmatrix} \boldsymbol{B} \begin{bmatrix} \boldsymbol{U}^{\mathrm{H}} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{U}^{\mathrm{H}} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{U}^{\mathrm{H}} \end{bmatrix} \tag{A12}$$

where the matrix $\boldsymbol{B}$ is (A13), shown at the bottom of the page. In a similar way, $\boldsymbol{R}_{H_{:,n-q}\widetilde{H}} = E\left\{H_{:,n-q}\widetilde{\boldsymbol{H}}^{\mathrm{H}}\right\}$ is given by

$$\boldsymbol{R}_{H_{:,n-q}\widetilde{H}} = \boldsymbol{U}\boldsymbol{A} \begin{bmatrix} \boldsymbol{U}^{\mathrm{H}} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{U}^{\mathrm{H}} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{U}^{\mathrm{H}} \end{bmatrix} \tag{A14}$$

where the matrix $\boldsymbol{A}$ is $\boldsymbol{A} = [r_t(-q)\Lambda \quad r_t(K-q)\Lambda \quad \cdots \quad r_t((M-1)K-q)\Lambda]$. Therefore, the MMSE estimator $\boldsymbol{C}_q$ can be written as

$$\boldsymbol{C}_q = \boldsymbol{U}c_q \begin{bmatrix} \boldsymbol{U}^{\mathrm{H}} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{U}^{\mathrm{H}} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{U}^{\mathrm{H}} \end{bmatrix} \tag{A15}$$

where $\boldsymbol{c}_q$ is defined by

$$\boldsymbol{c}_q = \boldsymbol{A}B. \tag{A16}$$

Note the P-by-MP matrix $\boldsymbol{c}_q$ has a following special form given in (A17), shown at the botoom of the page. The sequences $\{c^q_{i1}, c^q_{i2}, \ldots, c^q_{iM}\}$ represent the tap weights of the ith FIR filter in Fig. 2.

APPENDIX III

Given the N-by-P channel matrix $\boldsymbol{H}$ whose rank is $r \leq \min(N, P)$, there are two unitary matrices N-by-N $\boldsymbol{U}$ and P-by-P $\boldsymbol{V}$, such that we may write by the SVD

$$\boldsymbol{H} = \boldsymbol{U}\boldsymbol{\Lambda}V^{\mathrm{H}} \tag{A18}$$

where N-by-P matrix $\boldsymbol{\Lambda}$ is defined by

$$\boldsymbol{\Lambda} = \begin{bmatrix} \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}. \tag{A19}$$

The singular value λ's are ordered as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$. Note the matrices $\boldsymbol{H}H^{\mathrm{H}}$ and $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}$ may be written as

$$\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{\mathrm{H}}\boldsymbol{U}^{\mathrm{H}} \tag{A20}$$

and

$$\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H} = \boldsymbol{V}\boldsymbol{\Lambda}^{\mathrm{H}}\boldsymbol{\Lambda}V^{\mathrm{H}} \tag{A21}$$

respectively. The left-hand side of (21) becomes

$$\begin{aligned} \boldsymbol{G}^{\mathrm{H}} &= \boldsymbol{H}^{\mathrm{H}}(\boldsymbol{H}H^{\mathrm{H}} + \sigma^2\boldsymbol{I}_N)^{-1} \\ &= \boldsymbol{V}\boldsymbol{\Lambda}^{\mathrm{H}}\boldsymbol{U}^{\mathrm{H}}(\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{\mathrm{H}}\boldsymbol{U}^{\mathrm{H}} + \sigma^2\boldsymbol{I}_N)^{-1} \\ &= \boldsymbol{V}\boldsymbol{\Lambda}^{\mathrm{H}}(\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{\mathrm{H}} + \sigma^2\boldsymbol{I}_N)^{-1}U^{\mathrm{H}}. \end{aligned} \tag{A22}$$

The inverse matrix $(\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{\mathrm{H}} + \sigma^2\boldsymbol{I}_N)^{-1}$ is given by

$$\begin{aligned} &(\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{\mathrm{H}} + \sigma^2\boldsymbol{I}_N)^{-1} \\ &\quad = \mathrm{diag}\left(\underbrace{\frac{1}{|\lambda_1|^2 + \boldsymbol{\sigma}^2}, \ldots, \frac{1}{|\lambda_r|^2 + \boldsymbol{\sigma}^2}, \frac{1}{\boldsymbol{\sigma}^2}, \ldots, \frac{1}{\boldsymbol{\sigma}^2}}_{N}\right). \end{aligned} \tag{A23}$$

Hence, the P-by-N matrix $\boldsymbol{\Lambda}^{\mathrm{H}}(\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{\mathrm{H}} + \sigma^2\boldsymbol{I}_N)^{-1}$ is

$$\begin{aligned} &\boldsymbol{\Lambda}^{\mathrm{H}}(\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{\mathrm{H}} + \sigma^2\boldsymbol{I}_N)^{-1} \\ &\quad = \begin{bmatrix} \mathrm{diag}\left(\frac{\lambda_1^*}{|\lambda_1|^2+\sigma^2}, \ldots, \frac{\lambda_r^*}{|\lambda_r|^2+\sigma^2}\right) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}. \end{aligned} \tag{A24}$$

Correspondingly, we can express the matrix $\boldsymbol{G}^{\mathrm{H}}$ in the expanded form

$$\boldsymbol{G}^{\mathrm{H}} = \boldsymbol{V}\boldsymbol{\Lambda}^{\mathrm{H}}(\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{\mathrm{H}} + \sigma^2\boldsymbol{I}_N)^{-1}\boldsymbol{U}^{\mathrm{H}} = \sum_{i=1}^{r} \frac{\lambda_i^*}{|\lambda_i|^2 + \sigma^2}\boldsymbol{v}_i\boldsymbol{u}_i^{\mathrm{H}} \tag{A25}$$

where the column vector $\boldsymbol{v}_i$ and $\boldsymbol{u}_i$ are the ith column vector of the $\boldsymbol{V}$ and $\boldsymbol{U}$, respectively.

$$\boldsymbol{R}_{\widetilde{H}\widetilde{H}} = \begin{bmatrix} r_t(0)\boldsymbol{R}_f + \sigma^2\boldsymbol{I}_P & r_t(K)\boldsymbol{R}_f & \cdots & r_t((M-1)K)\boldsymbol{R}_f \\ r_t(-K)\boldsymbol{R}_f & r_t(0)\boldsymbol{R}_f + \sigma^2\boldsymbol{I}_P & \cdots & r_t((M-2)K)\boldsymbol{R}_f \\ \vdots & \vdots & \ddots & \vdots \\ r_t(-(M-1)K)\boldsymbol{R}_f & r_t(-(M-2)K)\boldsymbol{R}_f & \cdots & r_t(0)\boldsymbol{R}_f + \sigma^2\boldsymbol{I}_P \end{bmatrix} \tag{A10}$$

$$\boldsymbol{B} = \begin{bmatrix} r_t(0)\boldsymbol{\Lambda} + \sigma^2\boldsymbol{I}_P & r_t(K)\boldsymbol{\Lambda} & \cdots & r_t((M-1)K)\boldsymbol{\Lambda} \\ r_t(K)\boldsymbol{\Lambda} & r_t(0)L + \sigma^2\boldsymbol{I}_P & \cdots & r_t((M-2)K)\boldsymbol{\Lambda} \\ \vdots & \vdots & \ddots & \vdots \\ r_t((M-1)K)\boldsymbol{\Lambda} & r_t((M-2)K)\boldsymbol{\Lambda} & \cdots & r_t(0)\boldsymbol{\Lambda} + \sigma^2\boldsymbol{I}_P \end{bmatrix}^{-1} \tag{A13}$$

$$\boldsymbol{c}_q = [\mathrm{diag}\,(c^q_{11}, \ldots, c^q_{r1}, 0, \ldots, 0) \quad \mathrm{diag}\,(c^q_{12}, \ldots, c^q_{r2}, 0, \ldots, 0) \quad \ldots \quad \mathrm{diag}\,(c^q_{1M}, \ldots, c^q_{rM}, 0, \ldots, 0)] \tag{A17}$$

The right-hand side of (21) can be written as

$$\begin{aligned}\boldsymbol{G}^{\mathrm{H}} &= (\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H} + \sigma^2\boldsymbol{I}_P)^{-1}\boldsymbol{H}^{\mathrm{H}} \\ &= (\boldsymbol{V}\boldsymbol{\Lambda}^{\mathrm{H}}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{H}} + \sigma^2\boldsymbol{I}_P)^{-1}\boldsymbol{V}\boldsymbol{\Lambda}^{\mathrm{H}}\boldsymbol{U}^{\mathrm{H}} \\ &= \boldsymbol{V}(\boldsymbol{\Lambda}^{\mathrm{H}}\boldsymbol{\Lambda} + \sigma^2\boldsymbol{I}_P)^{-1}\boldsymbol{\Lambda}^{\mathrm{H}}\boldsymbol{U}^{\mathrm{H}}. \end{aligned} \tag{A26}$$

Similarly, one can show that the matrix $(\Lambda^{\mathrm{H}}\Lambda + \sigma^2\boldsymbol{I}_p)^{-1}\Lambda^{\mathrm{H}}$ is equal to

$$\begin{aligned}&(\boldsymbol{\Lambda}^{\mathrm{H}}\boldsymbol{\Lambda} + \sigma^2\boldsymbol{I}_P)^{-1}\boldsymbol{\Lambda}^{\mathrm{H}} \\ &\quad = \begin{bmatrix} \operatorname{diag}\left(\frac{\lambda_1^*}{|\lambda_1|^2+\sigma^2}, \cdots, \frac{\lambda_r^*}{|\lambda_r|^2+\sigma^2}\right) & 0 \\ 0 & 0 \end{bmatrix}\end{aligned} \tag{A27}$$

which is the exactly same as (A24). Therefore, the right-hand side of (21) becomes

$$\boldsymbol{G}^{\mathrm{H}} = V(\boldsymbol{\Lambda}^{\mathrm{H}}\boldsymbol{\Lambda} + \sigma^2\boldsymbol{I}_P)^{-1}\boldsymbol{\Lambda}^{\mathrm{H}}\boldsymbol{U}^{\mathrm{H}} = \sum_{i=1}^{r} \frac{\lambda_i^*}{|\lambda_i|^2 + \sigma^2}\boldsymbol{v}_i\boldsymbol{u}_i^{\mathrm{H}}. \tag{A28}$$

REFERENCES

[1] R. W. Chang, "Synthesis of band-limited orthogonal signals for multichannel data transmission," *Bell Syst. Tech. J.*, vol. 45, pp. 1775–1796, Dec. 1966.

[2] R. W. Chang and R. A. Gibby, "A theoretical study of performance of an orthogonal multiplexing data transmission scheme," *IEEE Trans. Commun.*, vol. COM-16, pp. 529–540, Aug. 1968.

[3] B. R. Saltzberg, "Performance of an efficient parallel data transmission system," *IEEE Trans. Commun.*, vol. COM-15, pp. 805–811, Dec. 1967.

[4] L. J. Cimini, "Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing," *IEEE Trans. Commun.*, vol. COM-33, pp. 665–675, July 1985.

[5] S. Hara and R. Prasad, "Overview of multicarrier CDMA," *IEEE Commun. Mag.*, vol. 35, pp. 126–133, Dec. 1997.

[6] S. Kaiser, "On the performance of different detection techniques for OFDM-CDMA in fading channels," in *IEEE ICC '95*, Jun. 1995, pp. 2059–2063.

[7] Y. (G.) Li, L. J. Cimini Jr, and N. R. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Trans. Commun.*, vol. 46, pp. 902–914, July 1998.

[8] O. Edfors, M. Sandell, J.-J. van de Beek, S. K. Wilson, and P. O. Borjesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Commun.*, vol. 46, pp. 931–939, July 1998.

[9] W. G. Jeon, K. H. Chang, and Y. S. Cho, "An equalization technique for orthogonal frequency-division multiplexing systems in time-variant multipath channels," *IEEE Trans. Commun.*, vol. 47, pp. 27–32, Jan. 1999.

[10] P. A. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun. Syst.*, vol. 11, pp. 360–393, Dec. 1963.

[11] M. C. Jeruchim, P. Balaban, and K. S. Shanmugan, *Simulation of Communication Systems*. New York: Plenum, 1992.

[12] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channel," *IEEE Trans. Veh. Technol.*, vol. 40, pp. 686–693, Nov. 1991.

[13] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Reading, MA: Addison-Wesley, 1991.

[14] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. IEEE ISSSE-98*, Sep. 1998, pp. 295–300.

[15] Y. -S Choi, "Capacity Improvement for Multicarrier Signals in Time- and Frequency-Selective Rayleigh Fading Channel: Channel Estimation, Detection and Carrier Synchronization," Ph. D. dissertation, Polytechnic University, Brooklyn, NY, 2000.

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

Yang-Seok Choi (S'91–M'01) received the B.S. degree from Korea University, Seoul, Korea, in 1990, the M.S.E.E. degree from Korea Advanced Institute of Science and Technology, Taejon, Korea, in 1992, and the Ph.D. degree from Polytechnic University, Brooklyn, NY in 2000, all in electrical engineering.

From 1992 to 1996, he was with Samsung Electronics, Corporation Ltd, Suwon, Korea, where he developed 32 QAM modem for HDTV and QPSK ASIC for DBS. In 2000, he joined National Semiconductor, East Brunswick, NJ, where he was involved in the development of W-CDMA. Since 2001, he has been with AT&T Laboratories–Research, Middletown, NJ. His research interests include OFDM, MC-CDMA, W-CDMA, blind equalizer, carrier/timing recovery, space–time coding, and capacity of time-varying multipath channel. He holds seven U.S. patents.

Peter J. Voltz (S'78–A'80–M'81) received the B.S. and M.S. degrees in electrical engineering and the Ph.D. degree from Polytechnic University, Brooklyn, NY, in 1980, 1981, and 1987, respectively.

From 1981 to 1985, he was employed full time at Hazeltine Corporation in Greenlawn, NY, where he was involved in the design, analysis, and simulation of bandwidth efficient and spread spectrum radio systems. He has been a faculty member at Polytechnic University since the fall of 1987, after completing his doctoral work. His research interests have included the study of adaptive algorithms and their convergence behavior, parameter identification techniques in controlled Markov Chains, and source localization in the underwater environment. More recently, he has been focusing on the general area of wireless communication networks, with specific interests including optimization of signal acquisition and tracking techniques in dynamic multipath environments and adaptive power control and hand-off issues in cellular CDMA networks. Since 1997, he has also been working in the area of multicarrier modulation techniques for mobile communication networks. He is currently on a leave of absence from Polytechnic University, serving as the Vice President of Technology at LayerOne Wireless Technology, Melville, NY, a new company working on future advanced wireless networks. His research and development interests include space-time processing, joint detection, adaptive modulation, and related areas.

Frank A. Cassara (S'65–M'70–SM'85) received the B.S.E.E. degree from Rutgers University, in 1966, and the M.S.E.E. and Ph.D.(EE) degrees from the Polytechnic Institute of Brooklyn in 1968 and 1971, respectively.

He has been a faculty member at Polytechnic University, Brooklyn, NY, since 1970 and is currently Associate Department Head and Professor of Electrical Engineering, teaching undergraduate and graduate courses in electronics, wireless communications, and electronic engineering laboratories. He has also been active in experimental research programs relating to adaptive interference cancellers, application of surface acoustic wave devices to communication receivers, and spread spectrum wireless information networks. He also serves as Director of Polytechnic's Wireless Communication Laboratory. He has published over 30 journal and conference proceedings papers and received numerous research and education grants from such agencies as the National Science Foundation, Army Research Office, Air Force Office of Scientific Research as well as from industry. In 1994 he developed Polytechnic's popular Wireless Communications Laboratory with the help of an NSF Instructructional Laboratory Equipment Grant and matching funds from several wireless companies. During the summers of 1995, 1998, 1999, and 2000 he conducted an NSF Undergraduate Faculty Enhancement Workshop at Polytechnic's Long Island campus on Wireless Communications for Electrical Engineering and Computer Engineering faculty from various colleges and universities in the U.S. to bring theoretical and practical knowledge relating to wireless communications to a national student population.

Dr. Cassara has received numerous awards for excellence in teaching.