Channel Estimation for Multicarrier Modulation Systems Using a Time-Frequency Polynomial Model

Xiaowen Wang, *Member, IEEE,* and K. J. Ray Liu, *Senior Member, IEEE*

***Abstract*—Channel estimation is a crucial aspect in the design of multicarrier modulation systems. In this work, we propose a channel estimation scheme based on polynomial approximation of the channel responses in both time and frequency domains. The proposed estimator is more robust to the variations of channel statistics. Our simulation shows that it has more than 5-dB improvement over the existing methods under practical channel conditions.**

***Index Terms*—Channel estimation, multicarrier modulation, polynomial expansion.**

I. Introduction

MULTICARRIER modulation (MCM) is an effective technique for broadband wireless communications [1]. It partitions the entire bandwidth into parallel independent subchannels to transmit parallel low-bit-rate data streams. Thus, MCM has a relative longer symbol duration which provides great immunity to intersymbol interference (ISI) and impulse noise. The independence among subchannels simplifies the design of the equalizer and provides an easy method for transmitter optimization. Since the channel information is required in both equalization and transmitter optimization, channel estimation plays an important role in MCM system design. Most channel estimation schemes try to exploit the correlation of the channel responses of subchannels to reduce the noise and improve the estimates, though the subchannels are considered to be independent in principle when performing signal detection. Minimum mean squared error (MMSE) estimation can be obtained if the channel correlation function of is known by using the singular value decomposition of the correlation matrix [2]. However, in practice, the correlation function is usually not known and the channel statistics may vary by time. Our goal is to design an estimation scheme under the condition that the channel statistics are not known or not completely known. One such scheme proposed in [2]–[4] assumes that the channel correlation matrix can be diagonalized by a Fourier transform. The assumption is true when we consider infinite samples of the channel responses. In practice, we can only have finite observations, which may cause severe leakage using this type of approach.

In this work, instead of finding the eigenbasis of the channel correlation matrix, we approximate the channel responses by a certain model basis and minimize the estimation error by controlling the model error and residual noise. The polynomial model can be used as such a model to approximate the fading multipath channel if it is viewed as a smoothly varying function [9]. The polynomial approximation is done in the time domain in [6], [7] and in the frequency domain in [5]. In this paper, we use the polynomial approximation in both the time and frequency domains. Therefore, the noise can be further suppressed because fewer coefficients need to be estimated. Comparing to the Fourier transform used in [2]–[4], the error caused by polynomial model approximation is bounded and hence is more robust to the channel statistics and system parameters.

Paper approved by G. M. Vitetta, the Editor for Equalization and Fading Channels of the IEEE Communications Society. Manuscript received November 1, 2000; revised August 8, 2001 and November 20, 2001. This paper was presented in part at the IEEE VTC'01/Fall, Atlantic City, NJ, October 2001 and at IEEE GLOBECOM'01, San Antonio, TX, November 2001.

X. Wang is with the Wireless Systems Research, Agere Systems, Murray Hill, NJ 07974 USA (e-mail: xiaowenw@agere.com).

K. J. R. Liu is with the Electrical and Computer Engineering Department, University of Maryland, College Park, MD 20742 USA.

Publisher Item Identifier 10.1109/TCOMM.2002.800816.

II. Time–Frequency Polynomial Channel Model for MCM Systems

The MCM system divides the whole bandwidth B_d into L subchannels and modulates a block of data onto a set of subcarriers of corresponding subchannels. In most MCM systems, the subchannels are divided evenly, and the bandwidth of the subchannels is $\Delta f = B_d/L$. Input data are first buffered to blocks and then divided into L bit streams. These bit streams are mapped to some complex constellation points $X_{i,k}, i = 0, \ldots, L-1$ at the kth block. The modulation is implemented by an L-point inverse discrete Fourier transform (IDFT). Then the modulated data are passed through a P/S converter to form serial data $x_{i,k}$. A cyclic prefix which is the copy of the last v samples of $x_{i,k}$'s is inserted before sending $x_{i,k}$'s to the channel. Now it follows that the symbol duration is L/B_d; however, the actual block duration is $T_f = (L + v/B_d)$ with a sampling rate B_d. For a system with $B_d = 800$ kHz, $L = 512$, and $v = 64$, the block duration is $T_f = 720$ μs. Such a system will be used in the rest of this paper.

At the receiver, the prefix part is discarded. The demodulation is performed by the discrete Fourier transform (DFT) operation. The demodulated data are the $Y_{i,k}$'s. If the cyclic prefix is sufficiently long, the interference between two MCM blocks is eliminated and the subchannels can be viewed as independent of each other, i.e.

$$Y_{i,k} = H_{i,k}X_{i,k} + N_{i,k} \tag{1}$$

where $H_{i,k}$ is the channel frequency response at $i\Delta f$ of the kth block and $N_{i,k}$ is the corresponding channel noise. $N_{i,k}$ is assumed to be a white Gaussian process with zero mean and variance σ^2.

Because of this simple relationship, only a one-tap equalizer is needed for each subchannel at the receiver, i.e., $\hat{X}_{i,k} = Y_{i,k}W_{i,k}$, where the equalizer coefficient $W_{i,k}$ is some function of $H_{i,k}$. For example, the zero-forcing equalizer is constructed as $W_{i,k} = 1/(H_{i,k})$. Then the decision is made upon $\hat{X}_{i,k}$. The problem for us is to estimate $H_{i,k}$.

In wireless broadband communications, the channel impulse response can be modeled as [8]

$$h(t,\tau)=\sum_i \gamma_i(t)\delta(\tau-\tau_i) \tag{2}$$

where $\gamma_i(t)$'s are independent narrow-band Gaussian processes with zero mean and variance p_i. All $\gamma_i(t)$'s have the same bandwidth which is defined as Doppler frequency f_D. (p_i,τ_i) defines the delay profile describing the channel dispersion which is also often characterized by the maximum delay $T_d \triangleq \max_i \tau_i$. Three types of delay profiles are used in this paper: TU, HT, and 2-ray. The TU and HT delay profiles both have six paths [8], while the 2-ray delay profile has two equal power paths. We also assume that the channel is normalized in our simulation, i.e., $\sum_i p_i = 1$.

The channel responses $H_{i,k}$'s are the samples of $H(t,f) = \int h(t,\tau)e^{-j2\pi\tau f}\,d\tau$, that is, $H_{i,k} = H(kT_f, i\Delta f)$. It is obvious that the Fourier transform of $H(t,f)$ is band-limited by f_D and T_d. Therefore, by discarding the high-frequency components out of the band, we can reduce the noise and improve the estimation. This is the idea used in [2]–[4]. However, the problem is that we only have finite samples of $H_{i,k}$ in a practical MCM system. The Fourier transform over these finite samples may suffer severe leakage, which degrades the performance dramatically.

On the other hand, the band-limited nature of the channel response suggests that the channel variation in the physical world is smooth in both the time and frequency domains. We know from the approximation theory [9] that such a smoothly varying function can be approximated by projecting to a finite set of basis functions. Moreover, since the MCM channel responses are located in a time–frequency plane, it is natural to project the responses over a time–frequency window $(2I+1)\Delta f \times (2K+1)T_f$ to a small set of polynomial basis functions around a center point (i_0, k_0), i.e.,

$$H_{i,k} = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} H_{i_0,k_0}(nm)(k-k_0)^m(i-i_0)^n + R_{MN},$$
$$\text{for } k_0-K \le k \le k_0+K;\, i_0-I \le i \le i_0+I \tag{3}$$

where M and N are the model orders for frequency and time domains, respectively, $H_{i_0,k_0}(nm) = (T_f^m \Delta f^n)/(m!n!)(\partial^m\partial^n H(t,f))/(\partial t^m \partial f^n)|_{t=k_0T_f, f=i_0\Delta f}$ and $R_{MN} = R_M + R_N - (((k-k_0)T_f)^M((i-i_0)\Delta f)^N)/(M!N!)(\partial^M \partial^N H(t,f))/(\partial t^M \partial f^N)|_{t=t', f=f'}, R_M = (((k-k_0)T_f)^M)/(M!)(\partial^M H(t,f))/(\partial t^M)|_{t=t'}$ and $R_N = (((i-i_0)\Delta f)^N/N!)(\partial^N H(t,f))/(\partial f^N)|_{f=f'}$ with $k_0T_f \le t' \le kT_f$ and $i_0\Delta f \le f' \le i\Delta f$.

The mean squared model error is then bounded by

$$\begin{aligned}E[\|R_{MN}\|^2] \le& \left(\frac{(KT_f)^M}{M!}\right)^2 \int_0^{f_D} (2\pi\xi)^M S_t(\xi)d\xi \\ &+ \left(\frac{(I\Delta f)^N}{N!}\right)^2 \int_0^{T_d} (2\pi\nu)^N S_f(\nu)d\nu \\ &+ \left(\frac{(KT_f)^M}{M!}\right)^2 \left(\frac{(I\Delta f)^N}{N!}\right)^2 \int_0^{f_D} \\ &\times \int_0^{T_d} (2\pi\xi)^M(2\pi\nu)^N S_t(\xi)S_f(\nu)\,d\xi\,d\nu \end{aligned} \tag{4}$$

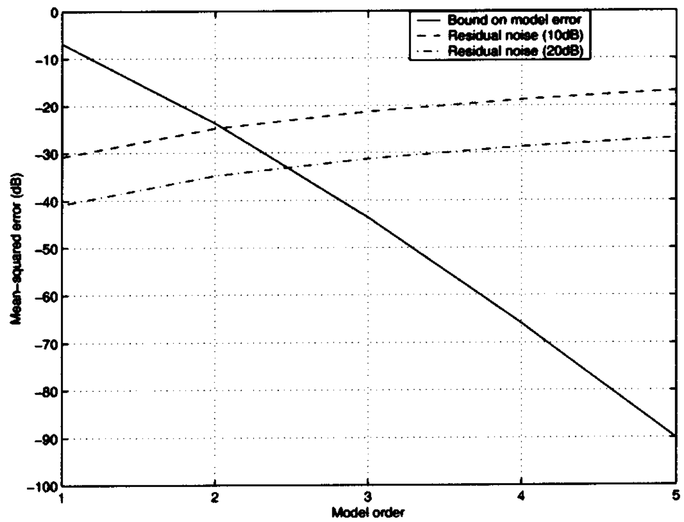


Fig. 1. Bound on the mean squared model error.

where $S_t(\xi)$ and $S_f(\nu)$ are the power spectra of the channel time- and frequency-domain correlation, respectively.

Without loss of generality, assuming $M = N$ and $v \lll L$ and using the multipath Rayleigh fading model, this bound can be derived as

$$\begin{aligned}E[\|R_{MM}\|^2] \le& \frac{2M!(2\pi KLf_D)^{2M}}{(2B_d)^{2M}(M!)^4} + \frac{(2\pi IB_dT_d)^{2M}}{L^{2M}(M!)^2} \\ &+ \frac{2M!(4\pi^2KIf_DT_d)^{2M}}{2^{2M}(M!)^6}.\end{aligned} \tag{5}$$

The first term in (5) is determined by Lf_D/B_d, while the second term is determined by B_dT_d/L. The third term is actually determined by f_DT_d and is much smaller than the first two terms if they are both smaller than one. To make the model error small, we can choose a larger model order M if $Lf_D/B_d < 1$ and $B_dT_d/L < 1$. However, the goal for modeling is to express the channel responses by a small number of model coefficients, which means that we want M to be small. The other way to reduce the above bound is to adjust the window dimensions K and I. For fixed f_DT_d, when L is large, the first term is dominating, then we should choose a smaller value of K to make the model error small. On the other hand, when L is small and the second term is dominating, then we should choose a smaller value of I. By carefully choosing the window dimensions, the time–frequency model error can be limited to a certain level once the Doppler frequency f_D, maximum delay T_d, and bandwidth B_d are fixed. It should be pointed out that, unlike using only a time- or frequency-domain model [5]–[7], the model error of the time–frequency model does not depend on the number of subchannels L.

Fig. 1 shows the upper bound of a mean squared model error with $f_DT_f = T_d\Delta f = 10^{-2}$ and window size $I = K = 5$ according to model order M. In this figure, we also show the residual noise for SNRs of 10 and 20 dB. It shows that the noise can be greatly reduced with a small penalty on model error. Moreover, such a model approximation does not need to know the actual channel correlation function.

III. CHANNEL ESTIMATION WITH A POLYNOMIAL MODEL

Suppose we have chosen the appropriate model orders and window dimensions such that

$$\mathbf{H}_{i_0,k_0} \simeq \mathbf{Q}_{M,N}(I,K)\mathbf{b}_{i_0,k_0} \tag{6}$$

where $\mathbf{H}_{i_0,k_0}$ and $\mathbf{Q}_{M,N}(I,K)$ are defined at the bottom of the page, with $q_{m,n}^{i,k} = i^n k^m$, for $i = -I,\ldots,0,\ldots,I, k = -K,\ldots,0,\ldots,K, m = 0,\ldots,M-1$ and $n = 0,\ldots,N-1$.

Then construct $\tilde{\mathbf{H}}_{i_0,k_0} = [\tilde{H}_{-I+i_0,-K+k_0} \cdots \tilde{H}_{-I+i_0,K+k_0} \cdots \tilde{H}_{I+i_0,-K+k_0} \cdots \tilde{H}_{I+i_0,K+k_0}]^T$ with $\tilde{H}_{i,k} = (Y_{i,k}/X_{i,k}) = H_{i,k} + (N_{i,k}/X_{i,k})$ as the temporary estimation. $X_{i,k}$'s can be obtained either from training or from the detected signal. Then

$$\tilde{\mathbf{H}}_{i_0,k_0} = \mathbf{H}_{i_0,k_0} + \mathbf{N}_{i_0,k_0} \simeq \mathbf{Q}_{M,N}(I,K)\mathbf{b}_{i_0,k_0} + \mathbf{N}_{i_0,k_0} \tag{7}$$

where $\mathbf{N}_{i_0,k_0}$ is defined at the bottom of the page.

Using least square methods, we can get the estimation of the coefficients of the polynomial basis from the temporary estimation

$$\hat{\mathbf{b}}_{i_0,k_0} = \mathbf{Q}_{M,N}^{\dagger}(I,K)\tilde{\mathbf{H}}_{i_0,k_0} \tag{8}$$

where $\mathbf{Q}_{M,N}^{\dagger}(I,K)$ is the pseudoinverse of $\mathbf{Q}_{M,N}(I,K)$.

The channel estimation then can be constructed as

$$\begin{aligned}\hat{H}_{i,k} &= \mathbf{q}_{M,N}(i-i_0,k-k_0)^T\hat{\mathbf{b}}_{i_0,k_0}\\ &= \mathbf{q}_{M,N}(i-i_0,k-k_0)^T\mathbf{Q}^{\dagger}(I,K)\tilde{\mathbf{H}}_{i_0,k_0}\end{aligned} \tag{9}$$

where $\mathbf{q}_{M,N}(i-i_0,k-k_0) = [q_{0,0}^{i-i_0,k-k_0} \cdots q_{0,N-1}^{i-i_0,k-k_0} \cdot q_{M-1,0}^{i-i_0,k-k_0} \cdots q_{M-1,N-1}^{i-i_0,k-k_0}]^T$. Usually we fix the value of $i-i_0$ and $k-k_0$, i.e., fix the point of estimation inside the window and slide the window to get all the estimations. Such an estimator can be viewed as a two-dimensional filtering process. Moreover, the polynomial basis has a symmetric property and a recursive algorithm can be derived to implement the filtering which reduces the computation complexity.

With an estimation point chosen to be at the center of the frequency domain window and an end point at the time domain window, assuming $E[\|X_{i,k}\|^2] = 1$, the estimation error from (9) becomes

$$\epsilon_{I,K} = E[\|H_{i_0,k_0} - \hat{H}_{i_0,k_0}\|^2] = \epsilon_h + \epsilon_n \tag{10}$$

where

$$\begin{aligned}\epsilon_h =& r_H(0,0) - E[H_{i_0,k_0}\mathbf{H}_{i_0,k_0}^T]\mathbf{Q}_{M,N}^{\dagger T}(I,K)\mathbf{q}_{M,N}(0,K)\\ &- \mathbf{q}_{M,N}^T(0,K)\mathbf{Q}_{M,N}^{\dagger}(I,K)E[\mathbf{H}_{i_0,k_0}H_{i_0,k_0}^*]\\ &+ \mathbf{q}_{M,N}^T(0,K)\mathbf{Q}_{M,N}^{\dagger}(I,K)E[\mathbf{H}_{i_0,k_0}\mathbf{H}_{i_0,k_0}^T]\\ &\cdot \mathbf{Q}_{M,N}^{\dagger T}(I,K)\mathbf{q}_{M,N}(0,K)\end{aligned}$$

is the model error and

$$\epsilon_n = \sigma^2\mathbf{q}_{M,N}^T(0,K)\mathbf{Q}_{M,N}^{\dagger}(I,K)\mathbf{Q}_{M,N}^{\dagger T}(I,K)\mathbf{q}_{M,N}(0,K)$$

is the residual noise.

The model error $\epsilon_h = 0$, if $\mathbf{Q}_{M,N}^{\dagger}(I,K)E[\mathbf{H}_{i_0,k_0}\mathbf{H}_{i_0,k_0}^T]\mathbf{Q}_{M,N}^{\dagger T}$ is a diagonal matrix. This can be realized using the eigenbasis of the channel correlation function. However, the statistics of the channel must be known which is difficult in practice and also difficult to implement. Then, for a model basis like the polynomial model, the model error increases while the residual noise decreases when the model order $M \times N$ becomes smaller or the window dimension $I \times K$ becomes larger. The tradeoff can be reached by adjusting the window dimensions and model orders to the channel statistics.

IV. SIMULATION RESULTS

The MCM system used in the simulations is the system introduced in Section II. Fig. 2 shows the comparison of mean-squared estimation errors of the channel estimates based on expansions in both the time and frequency domains with those based on expansion in either the time or frequency domain. We can see that the estimation error with both time- and frequency-domain expansions is about 3 dB less at an SNR of 10 dB compared to the frequency-domain expansion [5] and more than 7 dB less compared to the time-domain expansion [6].

Fig. 3 shows the estimation error under different delay profiles with a Doppler frequency of 40 Hz. Fig. 3(a) shows the es-

$$\mathbf{H}_{i_0,k_0} = [H_{-I+i_0,-K+k_0} \quad \cdots \quad H_{-I+i_0,K+k_0} \quad \cdots \quad H_{I+i_0,-K+k_0} \quad \cdots \quad H_{I+i_0,K+k_0}]^T$$

$$\mathbf{b}_{i_0,k_0} = [H_{i_0,k_0}(0,0) \quad \cdots \quad H_{i_0,k_0}(0,N-1) \quad \cdots \quad H_{i_0,k_0}(M-1,0) \quad \cdots \quad H_{i_0,k_0}(M-1,N-1)]^T$$

$$\mathbf{Q}_{M,N}(I,K) = \begin{bmatrix} q_{0,0}^{-I,-K} & \cdots & q_{0,N-1}^{-I,-K} & q_{1,0}^{-I,-K} & \cdots & q_{M-1,N-1}^{-I,-K} \\ \vdots & & \vdots & \vdots & & \vdots \\ q_{0,0}^{I,-K} & \cdots & q_{0,N-1}^{I,-K} & q_{1,0}^{I,-K} & \cdots & q_{M-1,N-1}^{I,-K} \\ \vdots & & \vdots & \vdots & & \vdots \\ q_{00}^{I,K} & \cdots & q_{0,N-1}^{I,K} & q_{1,0}^{I,K} & \cdots & q_{M-1,N-1}^{I,K} \end{bmatrix}$$

$$\mathbf{N}_{i_0,k_0} = \left[\frac{N_{-I+i_0,-K+k_0}}{X_{-I+i_0,-K+k_0}} \quad \cdots \quad \frac{N_{-I+i_0,K+k_0}}{X_{-I+i_0,K+k_0}} \quad \cdots \quad \frac{N_{I+i_0,-K+k_0}}{X_{I+i_0,-K+k_0}} \quad \cdots \quad \frac{N_{I+i_0,K+k_0}}{X_{I+i_0,K+k_0}}\right]^T$$

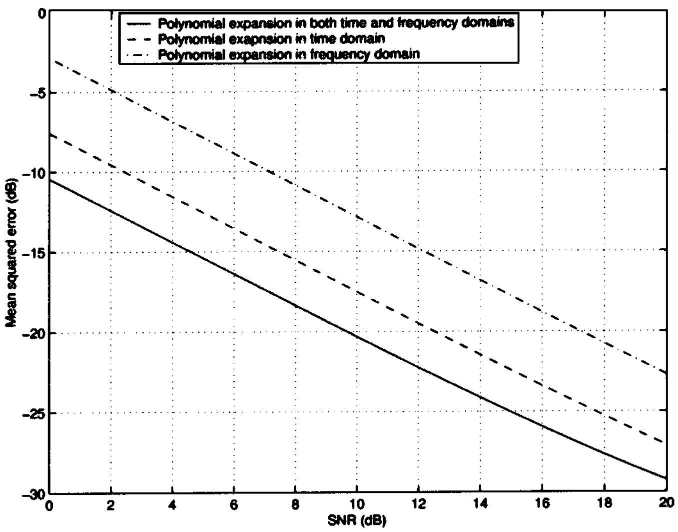
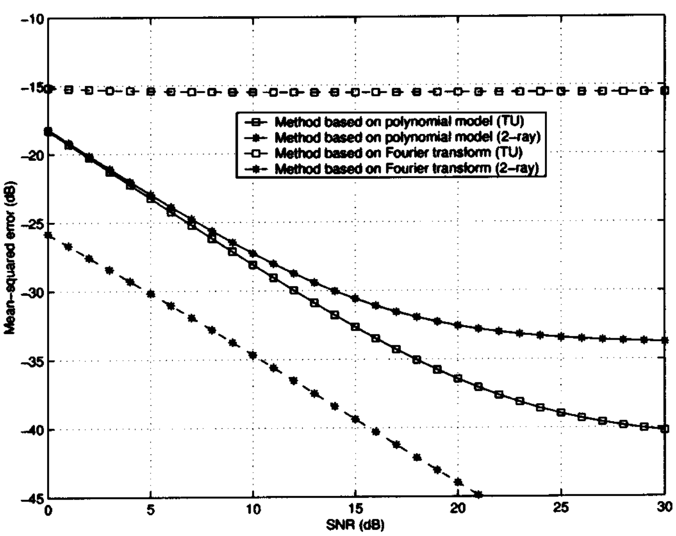
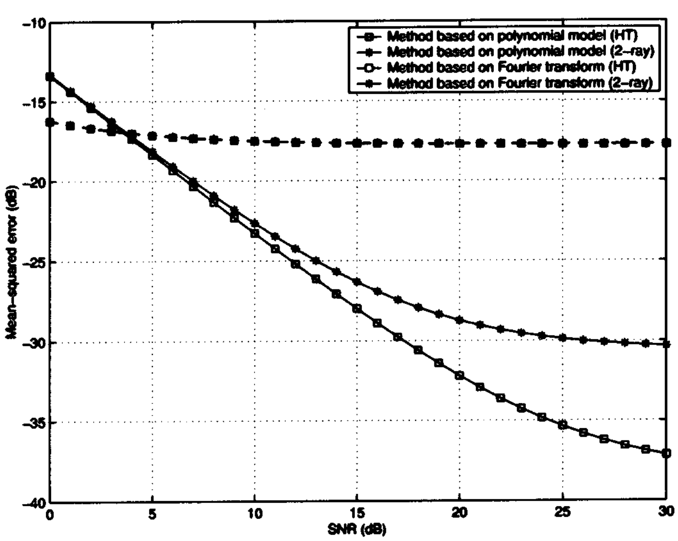


Fig. 2. Estimation error versus SNR (2-ray, $f_D = 30$ Hz and $T_d = 25\ \mu$s, $I \times K = 6 \times 6$ and $M \times N = 3 \times 3$).



(a)



(b)

Fig. 3. Estimation error versus SNR $(M \times N = 3 \times 3)$. (a) $I \times K = 40 \times 4$ and $T_d = 5\ \mu$s (b) $I \times K = 11 \times 5$ and $T_d = 17.2\ \mu$s.

timation error with a TU delay profile and a 2-ray delay profile with the same maximal delay as TU and (b) shows the estimation error with an HT delay profile and a 2-ray delay profile with the same maximal delay as HT. The results using the Fourier transform-based method of [4] are also shown for comparison. With a finite number of subchannels, all the delay paths of the channel have to be at the sampling instances of the system to avoid leakage, otherwise severe performance loss occurs. For a 2-ray channel with $T_d = 5\ \mu$s that is the maximal delay of TU, the method in [4] exhibits a better performance since the two delay paths at $T_d = 0$ and $T_d = 5\ \mu$s are both at the sampling instance of the system, and hence there is no leakage caused by Fourier transform. However, the leakage becomes large for TU or HT delay profiles because not all their delay paths are at the sampling instances and the performance of the Fourier transform-based method is greatly degraded. If we consider the Fourier transform as the approximation model basis, in those cases with leakage, the proposed polynomial model method has much less model error and therefore it has more than 5 dB gain over the Fourier transform-based method than the SNR larger than 10 dB and more robust to the channel statistics variation. There is only a small difference between the TU or HT profiles and its corresponding 2-ray channel with same T_d, respectively.

V. Conclusion

In this paper, we studied the channel estimation problem for the MCM system when the statistics of the multipath fading channel are not known or are partially known. A channel estimation approach based on a time–frequency polynomial model of the channel response is proposed. The method exploits the channel correlation in both the time and frequency domains. It is shown in simulation that the method is robust to different channel statistics.

References

[1] J. A. C. Bingham, “Multicarrier modulation for data transmission: An idea whose time has come,” *IEEE Commun. Mag.*, pp. 5–14, May 1990.

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

[3] V. Mignone and A. Morello, “CD3-OFDM: A novel demodulation scheme for fixed and mobile receivers,” *IEEE Trans. Commun.*, vol. 44, pp. 1141–1151, Sept. 1996.

[4] Y. Li, L. J. Cimini Jr., and N. R. Sollengerger, “Robust channel estimation for OFDM systems with rapid dispersive fading channels,” *IEEE Trans. Commun.*, vol. 46, pp. 902–915, July 1998.

[5] M. Luise, R. Reggiannini, and G. M. Vitetta, “Blind equalization/detection for OFDM signals over frequency-selective channels,” *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1568–1578, Oct. 1998.

[6] D. K. Borah and B. D. Hart, “A robust reciever structure for time-varying, frequency-flat Rayleigh fading channels,” *IEEE Trans. Commun.*, vol. 47, pp. 360–364, Mar. 1999.

[7] ——, “Frequency-selective fading channel estimation with a polynomial time-varying channel model,” *IEEE Trans. Commun.*, vol. 47, pp. 862–873, June 1999.

[8] W. C. Jakes, *Microwave Mobile Communications*. New York: Wiley, 1974.

[9] E. W. Cheney, *Introduction to Approximation Theory*. New York: McGraw-Hill, 1966.