A Derivation on the Equivalence Between Newton's Method and DF DFT-Based Method for Channel Estimation in OFDM Systems

Meng-Lin Ku and Chia-Chi Huang

***Abstract*—In this paper, we derive the decision-feedback (DF) discrete Fourier transform (DFT)-based channel estimation method from Newton's method for space-time block code (STBC)/ orthogonal frequency division multiplexing (OFDM) systems. Through our derivation, the equivalence between Newton's method and the DF DFT-based method is established. Computer simulations are also used to demonstrate the equivalence of the two methods in terms of BER and normalized square error (NSE) performance. Finally, the results presented in this paper also hold for conventional OFDM systems.**

***Index Terms*—Orthogonal frequency division multiplexing, space-time block code, discrete Fourier transform, channel estimation, Newton's method, decision-feedback.**

I. Introduction

The discrete Fourier transform (DFT)-based channel estimation method derived from the maximum likelihood (ML) criterion is originally proposed for orthogonal frequency division multiplexing (OFDM) systems with pilot preambles [1]–[5]. In order to save bandwidth and improve system performance, decision-feedback (DF) data symbols are usually exploited to track channel variations in subsequent OFDM data symbols, and this method is called DF DFT-based channel estimation [1]–[3]. However, the working principle of this empirical method has not been explored from the viewpoint of Newton's method in previous studies. This paper derives the DF DFT-based channel estimation via Newton's method for space-time block code (STBC)/OFDM systems. In this way, the equivalence between the two methods is established. Our results indicate that both methods can be implemented through the same four components: a least-square (LS) estimator, an inverse DFT (IDFT) matrix, a weighting matrix, and a DFT matrix, but with different connections. On one hand, the gradient vector in Newton's method can be found by calculating the difference between an estimated channel frequency response and an LS estimate, followed by the IDFT operation. On the other hand, the inverse of the Hessian matrix in Newton's method is just the weighting matrix operation in the DF DFT-based method.

Manuscript received June 1, 2007; revised June 3, 2008; accepted July 16, 2008. The associate editor coordinating the review of this paper and approving it for publication was X.-G. Xia. This work is supported in part by National Science Council under Grants NSC94-2220-E-009-033, NSC95-2220-E-009-010 and NSC96-2220-E-009-002.
The authors are with the Department of Communication Engineering, National Chiao Tung University, Hsinchu, 300, Taiwan (e-mail: willyku.cm87@nctu.edu.tw, huanggcc@faculty.nctu.edu.tw).
Digital Object Identifier 10.1109/T-WC.2008.070583

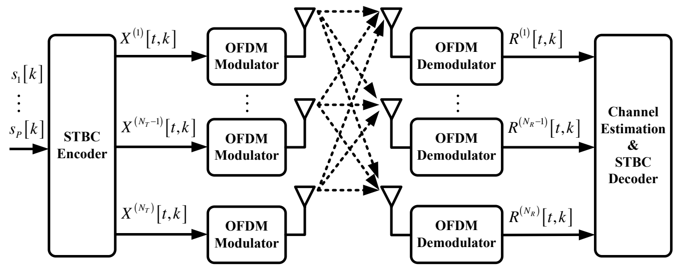

Fig. 1. STBC/OFDM system.

The rest of this paper is organized as follows. In Section II, we briefly describe an STBC/OFDM system. In Section III, a classical DF DFT-based channel estimation method is introduced, and a channel estimation method using the ML criterion is derived from Newton's method. The equivalence between the DF DFT-based method and Newton's method is then discussed and simulated in this section. Finally, some conclusions are drawn in Section IV.

II. STBC/OFDM Systems

Consider an STBC/OFDM system in Fig. 1 with N_T transmit and N_R receive antennas, employing K subcarriers among which M subcarriers are used to transmit data symbols and the other $K-M$ subcarriers are used as either a DC subcarrier or virtual subcarriers. Assume that the set of data subcarrier indices is denoted as $\mathbf{Q} \subseteq \{1,\ldots,K\}$. At subcarrier $k \in \mathbf{Q}$ and after symbol mapping, P modulated data symbols $\{s_1[k],\ldots,s_P[k]\}$ are encoded by an $N_T \times N_L$ STBC encoder $\mathbf{X}[k]$ to generate N_T signal sequences of length N_L, denoted by $\{X^{(i)}[1,k],\ldots,X^{(i)}[N_L,k]\}$, for $i=1,\ldots,N_T$ [6][7]. As a simple example, for a 2×2 Alamouti's STBC, we have $X^{(1)}[1,k]=s_1[k]$, $X^{(2)}[1,k]=s_2[k]$, $X^{(1)}[2,k]=-s_2^*[k]$, and $X^{(2)}[2,k]=s_1^*[k]$. Note that these signal sequences possess the orthogonal property, given by $\mathbf{X}^*[k]\mathbf{X}^T[k]=C[k]\mathbf{I}_{N_T}$, where $(\cdot)^*$ and $(\cdot)^T$ represent complex conjugate and transpose, respectively, $\mathbf{I}_N$ is an $N\times N$ identity matrix, and $C[k]=\sum_{t=1}^{N_L}|X^{(i)}[t,k]|^2$. After insertion of $K-M$ zeros for DC and virtual subcarriers, the STBC encoded data symbols $X^{(i)}[t,k]$ are modulated onto M subcarriers via a K-point IDFT unit to produce time domain samples, for $t=1,\ldots,N_L$ and $i=1,\ldots,N_T$. The time domain samples are then appended with cyclic prefix (CP) of length G and transmitted through N_T transmit antennas within the duration of N_L OFDM data symbols.

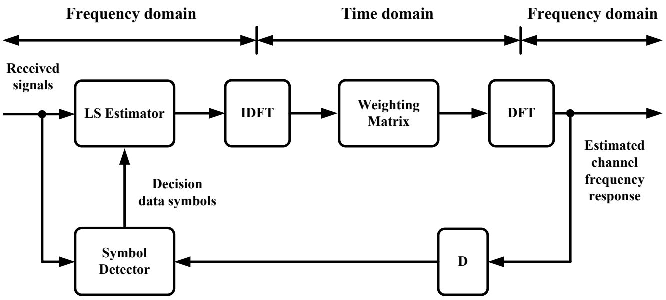

Fig. 2. The block diagram of the DF DFT-based channel estimation method. (D is a delay component.)

We assume that both timing and carrier frequency synchronization are perfect, and that the length of channel impulse response is always smaller than the length of the CP. Another assumption here is that the channel is quasi-static over the duration of a time slot, including N_L OFDM data symbols, but it varies from one time slot to another. Hence, at the output of the OFDM demodulator in Fig. 1, the N_L successively received OFDM data symbols at the jth receive antenna are given by

$$R^{(j)}\left[t,k\right]=\sum_{i=1}^{N_T}H^{(j,i)}\left[k\right]X^{(i)}\left[t,k\right]+Z^{(j)}\left[t,k\right] \quad (1)$$

for $t=1,\ldots,N_L$ and $k\in\mathbf{Q}$, where $H^{(j,i)}[k]$ is the channel frequency response for the $(j,i)th$ antenna pair, and $Z^{(j)}[t,k]$ is uncorrelated additive white Gaussian noise (AWGN) on the jth receive antenna with zero-mean and variance σ_Z^2.

III. DF DFT-Based Method and Newton's Method

A. DF DFT-Based Channel Estimation Method

As shown in Fig. 2, the block diagram of the DF DFT-based channel estimation method is composed of an LS estimator, an IDFT matrix, a weighting matrix, and a DFT matrix [2]–[5]. The LS estimator exploits DF data symbols to produce an LS estimate, which is a noisy estimation of channel frequency response. After taking the IDFT to transform the estimate to time domain, we can improve this estimate by using a weighting matrix which depends on the performance criterion chosen, either ML or minimum mean square error (MMSE) [3][5]. Finally, the enhanced estimate is transformed back to frequency domain to obtain a new estimate of channel frequency response.

B. Channel Estimation via Newton's Method

A parametric channel model $M^{(j,i)}[k]$ of the channel frequency response $H^{(j,i)}[k]$ is first formed by a summation of G complex sinusoids as follows:

$$M^{(j,i)}\left[k\right]=\sum_{l=1}^{G}\mu_l^{(j,i)}e^{-\jmath 2\pi(k-1)(l-1)/K} \quad (2)$$

where $\mu_l^{(j,i)}=\alpha_l^{(j,i)}+\jmath\beta_l^{(j,i)}$ is a complex fading gain to be tracked in subsequent time slots. From (1) and (2), the joint channel estimation and data detection problem can be formulated in an ML estimation framework as follows:

$$(\hat{\mathbf{s}},\hat{\mathbf{y}})=\arg\min_{\mathbf{s},\mathbf{y}}\sum_{j=1}^{N_R}\sum_{t=1}^{N_L}\sum_{k\in\boldsymbol{\Theta}}\Big|R^{(j)}\left[t,k\right]-\sum_{i=1}^{N_T}M^{(j,i)}\left[k\right]X^{(i)}\left[t,k\right]\Big|^2 \quad (3)$$

where $\boldsymbol{\Theta}=\{\Theta_1,\ldots,\Theta_{N_S}\}$ is a subset of $\mathbf{Q}$ over which we execute the summation, $\mathbf{s}$ denotes the data symbols which are STBC encoded and transmitted over subcarriers $\boldsymbol{\Theta}$, and N_S denotes the cardinality of $\boldsymbol{\Theta}$. In addition, we define $\boldsymbol{\mu}_I^{(j,i)}=[\alpha_1^{(j,i)},\ldots,\alpha_G^{(j,i)}]^T$, $\boldsymbol{\mu}_Q^{(j,i)}=[\beta_1^{(j,i)},\ldots,\beta_G^{(j,i)}]^T$, $\mathbf{y}^{(j,i)}=[\boldsymbol{\mu}_I^{(j,i)^T},\boldsymbol{\mu}_Q^{(j,i)^T}]^T$, $\mathbf{y}^{(j)}=[\mathbf{y}^{(j,1)^T},\ldots,\mathbf{y}^{(j,N_T)^T}]^T$, and $\mathbf{y}=[\mathbf{y}^{(1)^T},\ldots,\mathbf{y}^{(N_R)^T}]^T$. Because it is hard to solve (3) directly, we yield a simplified optimization problem by relaxing (3) as follows:

$$(\hat{\mathbf{s}},\hat{\mathbf{y}})=\arg\min_{\mathbf{y}}\min_{\mathbf{s}}\sum_{j=1}^{N_R}\sum_{t=1}^{N_L}\sum_{k\in\boldsymbol{\Theta}}\Big|R^{(j)}\left[t,k\right]-\sum_{i=1}^{N_T}M^{(j,i)}\left[k\right]X^{(i)}\left[t,k\right]\Big|^2 \quad (4)$$

Assuming that $M^{(j,i)}[k]$ is known, it is straightforward to solve the minimization problem with respect to $\mathbf{s}$ first by applying the STBC decoding algorithm [6][7], and we have

$$\begin{aligned}\hat{\mathbf{y}}&=\arg\min_{\mathbf{y}}\sum_{j=1}^{N_R}\sum_{t=1}^{N_L}\sum_{k\in\boldsymbol{\Theta}}\Big|R^{(j)}\left[t,k\right]-\sum_{i=1}^{N_T}M^{(j,i)}\left[k\right]\hat{X}^{(i)}\left[t,k\right]\Big|^2\\&\triangleq\arg\min_{\mathbf{y}}\sum_{j=1}^{N_R}\sum_{t=1}^{N_L}\sum_{k\in\boldsymbol{\Theta}}\Big|\Psi^{(j)}\left[t,k\right]\Big|^2\end{aligned} \quad (5)$$

where $\hat{X}^{(i)}[t,k]$ is the signal (corresponding to $X^{(i)}[t,k]$) obtained by re-encoding the decision symbols $\hat{s}_p[k]=\Phi(\tilde{s}_p[k])$, $\Phi(\cdot)$ is a symbol decision function, and $\tilde{s}_p[k]$ is the signal after diversity combining [6][7]. Notice that (5) might converge to local minima, leading to bit error rate (BER) performance loss, as compared with (3), particularly when the initial choice of $M^{(j,i)}[k]$ is not accurate enough. By rewriting (5), we have

$$\begin{aligned}\hat{\mathbf{y}}&=\arg\min_{\mathbf{y}}\sum_{j=1}^{N_R}\sum_{t=1}^{N_L}\sum_{k\in\boldsymbol{\Theta}}\Psi_I^{(j)^2}\left[t,k\right]+\Psi_Q^{(j)^2}\left[t,k\right]\\&\triangleq\arg\min_{\mathbf{y}}D\left(\mathbf{y}\right)\end{aligned} \quad (6)$$

where notations $\Upsilon_I(\cdot)$ and $\Upsilon_Q(\cdot)$ denote the real and imaginary part of the notation $\Upsilon(\cdot)$, respectively. For simplification, we drop the variable notation "$(\mathbf{y})$" in $D(\mathbf{y})$ hereafter except otherwise stated. Now we use Newton's method to find the minimum of (6), and the well-known iterative formula of Newton's method is provided in the following [8]:

$$\hat{\mathbf{y}}_v=\hat{\mathbf{y}}_{v-1}-\mathbf{g}_v \quad (7)$$

where v is the iteration index and $v=1,\ldots,V$, $\hat{\mathbf{y}}_v$ is the estimated channel state information (CSI) obtained at the vth

iteration, $\mathbf{g}_v$ is a search vector associated with $\mathbf{g} = \mathbf{E}^{-1}\mathbf{q}$ at $\mathbf{y} = \hat{\mathbf{y}}_{v-1}$ in which $\mathbf{E}$ and $\mathbf{q}$ are the Hessian matrix and the gradient vector of D, respectively, and $(\cdot)^{-1}$ represents the matrix inverse. Thus, the uth entry of $\mathbf{q}$ is calculated as

$$(\mathbf{q})_u \triangleq \frac{\partial D}{\partial (\mathbf{y})_u} = 2\sum_{j=1}^{N_R}\sum_{t=1}^{N_L}\sum_{k\in\boldsymbol{\Theta}} \Psi_I^{(j)}[t,k]\frac{\partial \Psi_I^{(j)}[t,k]}{\partial (\mathbf{y})_u} + \Psi_Q^{(j)}[t,k]\frac{\partial \Psi_Q^{(j)}[t,k]}{\partial (\mathbf{y})_u} \quad (8)$$

where $(\mathbf{y})_u$ is the uth entry of $\mathbf{y}$. The partial derivative of $\partial\Psi_I^{(j)}[t,k]/\partial(\mathbf{y})_u$ and $\partial\Psi_Q^{(j)}[t,k]/\partial(\mathbf{y})_u$ can be derived in the following way. First, we assume that the probabilities of $\tilde{s}_{p,I}[k] = 0$ or $\tilde{s}_{p,Q}[k] = 0$ are zero; thus, it is reasonable to take the terms involving the partial derivative of the function $\Phi(\cdot)$ as zero. Since the variable $(\mathbf{y})_u$ in $\mathbf{y}$ is either $\alpha_l^{(j,i)}$ or $\beta_l^{(j,i)}$, straightforward calculation using (6) shows that for $j = j'$, we have

$$\frac{\partial \Psi_I^{(j')}[t,k]}{\partial \alpha_l^{(j,i)}} = -\cos\left(\frac{2\pi(k-1)(l-1)}{K}\right)\hat{X}_I^{(i)}[t,k] - \sin\left(\frac{2\pi(k-1)(l-1)}{K}\right)\hat{X}_Q^{(i)}[t,k] \quad (9)$$

$$\frac{\partial \Psi_Q^{(j')}[t,k]}{\partial \alpha_l^{(j,i)}} = -\cos\left(\frac{2\pi(k-1)(l-1)}{K}\right)\hat{X}_Q^{(i)}[t,k] + \sin\left(\frac{2\pi(k-1)(l-1)}{K}\right)\hat{X}_I^{(i)}[t,k] \quad (10)$$

$$\frac{\partial \Psi_I^{(j')}[t,k]}{\partial \beta_l^{(j,i)}} = -\sin\left(\frac{2\pi(k-1)(l-1)}{K}\right)\hat{X}_I^{(i)}[t,k] + \cos\left(\frac{2\pi(k-1)(l-1)}{K}\right)\hat{X}_Q^{(i)}[t,k] \quad (11)$$

$$\frac{\partial \Psi_Q^{(j')}[t,k]}{\partial \beta_l^{(j,i)}} = -\sin\left(\frac{2\pi(k-1)(l-1)}{K}\right)\hat{X}_Q^{(i)}[t,k] - \cos\left(\frac{2\pi(k-1)(l-1)}{K}\right)\hat{X}_I^{(i)}[t,k] \quad (12)$$

Otherwise, i.e. if $j \neq j'$, we have

$$\frac{\partial \Psi_I^{(j')}[t,k]}{\partial \alpha_l^{(j,i)}} = \frac{\partial \Psi_Q^{(j')}[t,k]}{\partial \alpha_l^{(j,i)}} = \frac{\partial \Psi_I^{(j')}[t,k]}{\partial \beta_l^{(j,i)}} = \frac{\partial \Psi_Q^{(j')}[t,k]}{\partial \beta_l^{(j,i)}} = 0 \quad (13)$$

Next, we compute the $(m,u)th$ entry of $\mathbf{E}$ as

$$(\mathbf{E})_{m,u} \triangleq \frac{\partial^2 D}{\partial(\mathbf{y})_m\partial(\mathbf{y})_u} = 2\sum_{j=1}^{N_R}\sum_{t=1}^{N_L}\sum_{k\in\boldsymbol{\Theta}} \frac{\partial \Psi_I^{(j)}[t,k]}{\partial(\mathbf{y})_m}\frac{\partial \Psi_I^{(j)}[t,k]}{\partial(\mathbf{y})_u} + \frac{\partial \Psi_Q^{(j)}[t,k]}{\partial(\mathbf{y})_m}\frac{\partial \Psi_Q^{(j)}[t,k]}{\partial(\mathbf{y})_u} \quad (14)$$

where the terms involving the second derivative of $\Psi_I^{(j)}[t,k]$ (or $\Psi_Q^{(j)}[t,k]$) are all equal to zero. Since the variable in $\mathbf{y}$ is either $\alpha_l^{(j,i)}$ or $\beta_l^{(j,i)}$, the calculation of $\partial^2 D/\partial(\mathbf{y})_m\partial(\mathbf{y})_u$ is equivalent to finding $\partial^2 D/\partial\alpha_l^{(j,i)}\partial\alpha_{l'}^{(j',i')}$, $\partial^2 D/\partial\beta_l^{(j,i)}\partial\beta_{l'}^{(j',i')}$, $\partial^2 D/\partial\alpha_l^{(j,i)}\partial\beta_{l'}^{(j',i')}$, and $\partial^2 D/\partial\beta_l^{(j,i)}\partial\alpha_{l'}^{(j',i')}$in turn. By using (9)–(14) and the orthogonal property of STBC, as described in Section II, it follows that

$$\frac{\partial^2 D}{\partial\alpha_l^{(j,i)}\partial\alpha_{l'}^{(j',i')}} = \frac{\partial^2 D}{\partial\beta_l^{(j,i)}\partial\beta_{l'}^{(j',i')}} = \begin{cases} 0, \text{ if } i \neq i' \text{ or } j \neq j' \\ 2\sum\limits_{k\in\boldsymbol{\Theta}} \hat{C}[k]\cos\left(\frac{2\pi(k-1)(l-l')}{K}\right), \text{ o.w.} \end{cases} \quad (15)$$

$$\frac{-\partial^2 D}{\partial\alpha_l^{(j,i)}\partial\beta_{l'}^{(j',i')}} = \frac{\partial^2 D}{\partial\beta_l^{(j,i)}\partial\alpha_{l'}^{(j',i')}} = \begin{cases} 0, \text{ if } i \neq i' \text{ or } j \neq j' \\ 2\sum\limits_{k\in\boldsymbol{\Theta}} \hat{C}[k]\sin\left(\frac{2\pi(k-1)(l-l')}{K}\right), \text{ o.w.} \end{cases} \quad (16)$$

where $\hat{C}[k] = \sum_{t=1}^{N_L}|\hat{X}^{(i)}[t,k]|^2$. According to (14)–(16), we can make two observations. One is that the matrix $\mathbf{E}$ is related not only to the multipath delay l but also to the estimate of the total transmitted signal energy $\hat{C}[k]$ at the kth subcarrier for each transmit antenna. The other observation is that the matrix $\mathbf{E}$ is a block diagonal matrix. Owing to the second observation, the iterative channel estimation method in (7) can be further simplified to [1]

$$\hat{\mathbf{y}}_v^{(j,i)} = \hat{\mathbf{y}}_{v-1}^{(j,i)} - \mathbf{g}_v^{(j,i)} \quad (17)$$

where $\mathbf{g}_v^{(j,i)}$ is obtained by computing $\mathbf{g}^{(j,i)} = \mathbf{E}^{(j,i)^{-1}}\mathbf{q}^{(j,i)}$ at $\mathbf{y}^{(j,i)} = \hat{\mathbf{y}}_{v-1}^{(j,i)}$ in which $\mathbf{E}^{(j,i)}$ is the truncated matrix obtained from the $((j-1)N_T+i)th$ diagonal block of $\mathbf{E}$, and $\mathbf{q}^{(j,i)}$ is the truncated vector merely containing the partial derivate of $\partial D/\partial\alpha_l^{(j,i)}$ and $\partial D/\partial\beta_l^{(j,i)}$, for $i = 1,\ldots,N_T$ and $j = 1,\ldots,N_R$.

C. Equivalence between Newton's Method and DF DFT-Based Method

We now turn our attention to deriving the equivalence between Newton's method and the DF DFT-based channel estimation method. By using (8)–(12) and defining $\tilde{\mathbf{q}}^{(j,i)} =$

[1]It is noted that in this paper, $\mathbf{E}^{(j,i)}$ is a constant matrix for all antenna pairs. This is also the case for $\mathbf{F}^{(j,i)}$ and $\tilde{\mathbf{E}}^{(j,i)}$, but in practice, these matrices are specific for transceiver antenna pairs, depending upon path delays of the corresponding channel.

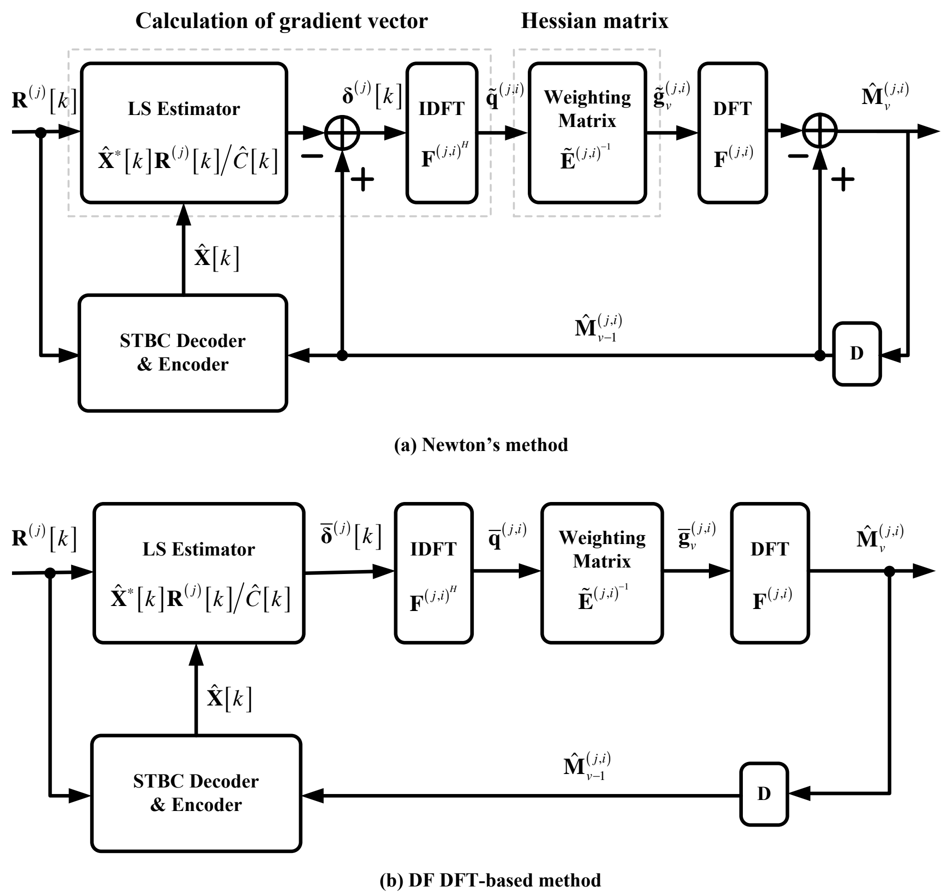

Fig. 3. Equivalence between (a) Newton's method and (b) the DF DFT-based method.

$\partial D/\partial \boldsymbol{\mu}_I^{(j,i)} + \jmath \partial D/\partial \boldsymbol{\mu}_Q^{(j,i)}$, the gradient vector $\mathbf{q}^{(j,i)}$ in (17) is rewritten in a complex vector form as follows [1]

$$\tilde{\mathbf{q}}^{(j,i)} = \mathbf{F}^{(j,i)^H} \boldsymbol{\Delta}^{(j,i)} \tag{18}$$

where $\boldsymbol{\Delta}^{(j,i)} = [\Delta^{(j,i)}[\Theta_1], \ldots, \Delta^{(j,i)}[\Theta_{N_S}]]^T$, each element of which is calculated by

$$\Delta^{(j,i)}[k] = -2 \sum_{t=1}^{N_L} \Psi^{(j)}[t,k] \hat{X}^{(i)^*}[t,k] \tag{19}$$

Moreover, $\mathbf{F}^{(j,i)}$ is an $N_S \times G$ truncated DFT matrix, with the $(m,l)th$ element given by $\exp\{-\jmath 2\pi(\Theta_m - 1)(l-1)/K\}$, and $(\cdot)^H$ is the Hermitian matrix of $(\cdot)$. By substituting $\Psi^{(j)}[t,k]$ of (5) into $\Delta^{(j,i)}[k]$ and applying the orthogonal property as described in Section II, a more meaningful expression is provided by rewriting (19) in a column vector form:

$$\begin{aligned}\boldsymbol{\delta}^{(j)}[k] &= \left[\Delta^{(j,1)}[k], \ldots, \Delta^{(j,N_T)}[k]\right]^T \\ &= 2\hat{\mathbf{X}}^*[k]\left(\hat{\mathbf{X}}^T[k]\mathbf{M}^{(j)}[k] - \mathbf{R}^{(j)}[k]\right) \\ &= 2\left(\hat{C}[k]\mathbf{M}^{(j)}[k] - \hat{\mathbf{X}}^*[k]\mathbf{R}^{(j)}[k]\right) \end{aligned} \tag{20}$$

for $k \in \boldsymbol{\Theta}$, where $\mathbf{R}^{(j)}[k] = [R^{(j)}[1,k], \ldots, R^{(j)}[N_L,k]]^T$, $\mathbf{M}^{(j)}[k] = [M^{(j,1)}[k], \ldots, M^{(j,N_T)}[k]]^T$, and $\hat{\mathbf{X}}[k]$ is the re-encoded STBC matrix with $\hat{X}^{(i)}[t,k]$ as its element. Here, we observe that $\hat{\mathbf{X}}^*[k]\mathbf{R}^{(j)}[k]$ is an LS estimate for $[H^{(j,1)}[k], \ldots, H^{(j,N_T)}[k]]^T$, and that $\boldsymbol{\delta}^{(j)}[k]$ represents the difference between the two channel frequency responses, $\hat{C}[k]\mathbf{M}^{(j)}[k]$ and $\hat{\mathbf{X}}^*[k]\mathbf{R}^{(j)}[k]$. From (18) and rewriting $\mathbf{E}^{(j,i)}$ in a complex matrix, $\tilde{\mathbf{E}}^{(j,i)}$, by using Appendix A, we have a complex-form representation of (17):

$$\hat{\boldsymbol{\mu}}_v^{(j,i)} = \hat{\boldsymbol{\mu}}_{v-1}^{(j,i)} - \tilde{\mathbf{g}}_v^{(j,i)} \tag{21}$$

where $\hat{\boldsymbol{\mu}}_v^{(j,i)}$ and $\tilde{\mathbf{g}}_v^{(j,i)}$ are the calculation associated with $\boldsymbol{\mu}^{(j,i)} = \boldsymbol{\mu}_I^{(j,i)} + \jmath \boldsymbol{\mu}_Q^{(j,i)}$ and $\tilde{\mathbf{g}}^{(j,i)} = \tilde{\mathbf{E}}^{(j,i)^{-1}} \tilde{\mathbf{q}}^{(j,i)}$, respectively, at $\boldsymbol{\mu}^{(j,i)} = \hat{\boldsymbol{\mu}}_{v-1}^{(j,i)}$ in which the $(l,l')th$ entry of $\tilde{\mathbf{E}}^{(j,i)}$ is given by $2\sum_{k\in\boldsymbol{\Theta}} \hat{C}[k] \exp\{\jmath 2\pi(k-1)(l-l')/K\}$. As shown in Appendix B, the matrix $\tilde{\mathbf{E}}^{(j,i)^{-1}}$ in effect acts as a path decorrelator to decorrelate inter-path interference. It is desirable for the path decorrelator to be independent of $\hat{C}[k]$; therefore, $\tilde{\mathbf{E}}^{(j,i)^{-1}}$ only needs to be calculated once in each OFDM frame, containing several time slots. One way to achieve this is to normalize $\boldsymbol{\delta}^{(j)}[k]$ in (20) by $2\hat{C}[k]$ and to modify $\tilde{\mathbf{E}}^{(j,i)}$ as follows [1]

$$\boldsymbol{\delta}^{(j)}[k] = \mathbf{M}^{(j)}[k] - \hat{\mathbf{X}}^*[k]\mathbf{R}^{(j)}[k]/\hat{C}[k] \tag{22}$$

$$\left(\tilde{\mathbf{E}}^{(j,i)}\right)_{l,l'} = \sum_{k\in\boldsymbol{\Theta}} e^{\jmath 2\pi(k-1)(l-l')/K} \tag{23}$$

To be precise, $\tilde{\mathbf{E}}^{(j,i)}$ in (23) can be equivalently expressed as $\mathbf{F}^{(j,i)^H}\mathbf{F}^{(j,i)}$. Finally, a truncated DFT matrix is applied to

TABLE I
SIMULATION PARAMETERS.

FFT size (K)	256
Number of data subcarriers (M)	200
Number of data subcarriers used (N_S)	200
Length of CP (G)	64
Modulation	QPSK
Number of transmit antennas (N_T)	2
Number of receive antennas (N_R)	1
Channel power profiles	ITU-Veh. A [9] and Jakes model [10]
Channel delay profiles	$0 \sim 63$ (Uniform distribution)
Maximum Doppler frequency (f_d)	0.01, 0.05

(21) to extrapolate the overall channel frequency response as follows

$$\hat{\mathbf{M}}_v^{(j,i)} = \hat{\mathbf{M}}_{v-1}^{(j,i)} - \mathbf{F}^{(j,i)}\tilde{\mathbf{g}}_v^{(j,i)} \tag{24}$$

where $\hat{\mathbf{M}}_v^{(j,i)} = \mathbf{F}^{(j,i)}\hat{\boldsymbol{\mu}}_v^{(j,i)}$ is the estimated channel frequency response at the vth iteration, with respect to the channel frequency response $\mathbf{M}^{(j,i)} = [M^{(j,i)}[\Theta_1], \ldots, M^{(j,i)}[\Theta_{N_S}]]^T$. It is clear that $\hat{\mathbf{M}}_{v-1}^{(j,i)}$ belongs to the subspace spanned by $\mathbf{F}^{(j,i)}$, and $\mathbf{F}^{(j,i)}\tilde{\mathbf{E}}^{(j,i)^{-1}}\mathbf{F}^{(j,i)^H}$ is an orthogonal projection onto this subspace. From matrix theory, these two observations imply that

$$\begin{aligned}
&\mathbf{F}^{(j,i)}\tilde{\mathbf{E}}^{(j,i)^{-1}}\mathbf{F}^{(j,i)^H}\hat{\mathbf{M}}_{v-1}^{(j,i)} \\
&= \mathbf{F}^{(j,i)}\tilde{\mathbf{E}}^{(j,i)^{-1}}\mathbf{F}^{(j,i)^H}\mathbf{F}^{(j,i)}\hat{\boldsymbol{\mu}}_{v-1}^{(j,i)} \\
&= \mathbf{F}^{(j,i)}\hat{\boldsymbol{\mu}}_{v-1}^{(j,i)} \\
&= \hat{\mathbf{M}}_{v-1}^{(j,i)}
\end{aligned} \tag{25}$$

Hence, the vector $\hat{\mathbf{M}}_{v-1}^{(j,i)} - \mathbf{F}^{(j,i)}\tilde{\mathbf{E}}^{(j,i)^{-1}}\mathbf{F}^{(j,i)^H}\hat{\mathbf{M}}_{v-1}^{(j,i)}$ implicitly contained in the right side of (24) is zero. As a result, (24) is reduced to the DF DFT-based channel estimation method and the equivalence can be expressed as [2]

$$\hat{\mathbf{M}}_v^{(j,i)} = \mathbf{F}^{(j,i)}\bar{\mathbf{g}}_v^{(j,i)} \tag{26a}$$

$$\bar{\mathbf{g}}^{(j,i)} = \tilde{\mathbf{E}}^{(j,i)^{-1}}\bar{\mathbf{q}}^{(j,i)} \tag{26b}$$

$$\bar{\mathbf{q}}^{(j,i)} = \mathbf{F}^{(j,i)^H}\bar{\boldsymbol{\Delta}}^{(j,i)} \tag{26c}$$

$$\bar{\boldsymbol{\delta}}^{(j)}[k] = \hat{\mathbf{X}}^*[k]\,\mathbf{R}^{(j)}[k]\,/\hat{C}[k]\,,\ \text{for } k \in \boldsymbol{\Theta} \tag{26d}$$

where we define $\bar{\boldsymbol{\Delta}}^{(j,i)} = [\bar{\Delta}^{(j,i)}[\Theta_1], \ldots, \bar{\Delta}^{(j,i)}[\Theta_{N_S}]]^T$ and $\bar{\boldsymbol{\delta}}^{(j)}[k] = [\bar{\Delta}^{(j,1)}[k], \ldots, \bar{\Delta}^{(j,N_T)}[k]]^T$, and $\bar{\mathbf{g}}_v^{(j,i)}$ is the calculation with respect to $\bar{\mathbf{g}}^{(j,i)}$ at $\mathbf{M}^{(j,i)} = \hat{\mathbf{M}}_{v-1}^{(j,i)}$.

The above derivation clearly establishes the mathematical equivalence between Newton's method of (24) shown in Fig. 3a and the DF DFT-based method of (26) shown in Fig. 3b. Our results indicate that both Newton's method and the DF DFT-based method for channel estimation in STBC/OFDM systems can be implemented through four components: an LS estimator, an IDFT matrix, a weighting matrix, and a DFT matrix. According to (18) and (22), we can also observe that the process of calculating the difference between the estimated channel frequency response $\mathbf{M}^{(j)}[k]$ and the LS estimate $\hat{\mathbf{X}}^*[k]\mathbf{R}^{(j)}[k]/\hat{C}[k]$, followed by the IDFT matrix $\mathbf{F}^{(j,i)^H}$, is equivalent to forming the gradient vector in Newton's method.

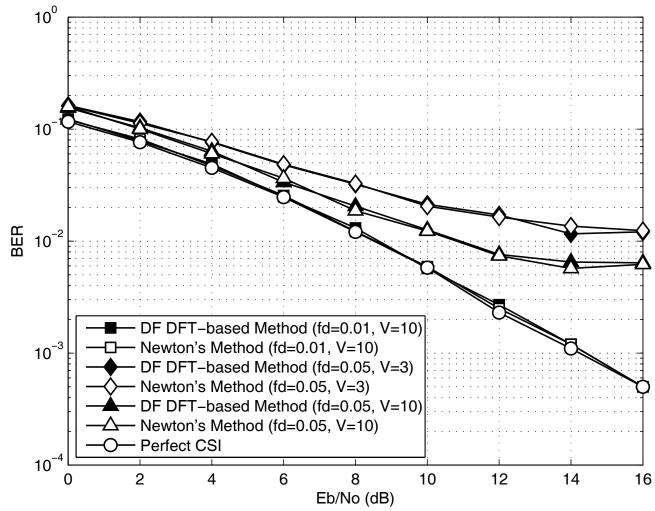

Fig. 4. BER performance of the two methods.

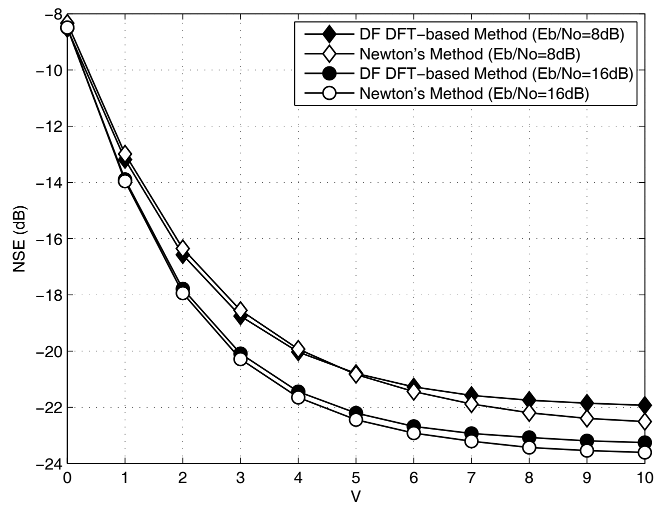

Fig. 5. NSE performance of the two methods. ($f_d = 0.05$).

Moreover, the weighting matrix $\tilde{\mathbf{E}}^{(j,i)^{-1}}$ in (26) is in fact the inverse of the Hessian matrix in Newton's method as observed in (21) and (23).

D. Simulation and Verification of the Equivalence between Two Methods

The equivalence of the two methods is also verified by simulation, and the results are given in Fig. 4 and Fig. 5, with parameters listed in Table I. We assume that the CSI in the previous time slot is known and utilized to initialize channel estimators. The parameter f_d denotes the maximum Doppler frequency, normalized to subcarrier spacing. The BER curve for ideal CSI is also provided for the purpose of calibration. As observed in the two figures, the equivalence of the two methods is demonstrated in terms of the performance of BER and normalized square error (NSE) between true and estimated CSI.

[2]Thanks to the orthogonal property of STBC, extending the results in single-input single-output/OFDM systems [2]–[4] to the STBC/OFDM systems should be straightforward.

IV. Conclusions

In this paper, we present a derivation on the equivalence between Newton's method and the DF DFT-based method for channel estimation in STBC/OFDM systems. The results could provide useful insights for the development of new algorithms. For example, extending the DF DFT-based method to the Levenberg-Marquardt method is quite simple through this equivalence, which is particularly helpful when the inverse for the weighting matrix does not exist [8]. As another example, a few pilot tones can be applied to form a gradient vector at the first iteration by using (21) and to help the DF DFT-based method jump out of local minimum, thus improving the BER performance in fast fading channels [11]. Finally, it is worth mentioning that the derivation and the relationships explored in this paper are also valid for conventional OFDM systems since they are only simplified cases of the systems discussed in this paper.

Appendix A.

Define $\mathbf{g}^{(j,i)} = [\mathbf{g}_1^T, \mathbf{g}_2^T]^T$ and $\mathbf{q}^{(j,i)} = [\mathbf{q}_1^T, \mathbf{q}_2^T]^T$, where $\mathbf{g}_1$, $\mathbf{g}_2$, $\mathbf{q}_1$, and $\mathbf{q}_2$ are of size $G \times 1$. From (15)–(17), we convert $\mathbf{g}^{(j,i)} = \mathbf{E}^{(j,i)^{-1}}\mathbf{q}^{(j,i)}$ into block matrix representation as follows

$$\begin{bmatrix}\mathbf{g}_1\\ \mathbf{g}_2\end{bmatrix} = \begin{bmatrix}\mathbf{A} & -\mathbf{B}\\ \mathbf{B} & \mathbf{A}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{q}_1\\ \mathbf{q}_2\end{bmatrix} \tag{A.1}$$

where $\mathbf{A}$ and $\mathbf{B}$ are $G \times G$ sub-matrices of $\mathbf{E}^{(j,i)}$. It is clear that the matrix $\mathbf{E}^{(j,i)^{-1}}$ holds the same structure as the matrix $\mathbf{E}^{(j,i)}$, given by

$$\mathbf{E}^{(j,i)^{-1}} = \begin{bmatrix}\mathbf{C} & -\mathbf{D}\\ \mathbf{D} & \mathbf{C}\end{bmatrix} \tag{A.2}$$

where $\mathbf{AC} - \mathbf{BD} = \mathbf{I}_G$, $\mathbf{BC} + \mathbf{AD} = \mathbf{0}_G$, and $\mathbf{0}_G$ is a zero matrix of size $G \times G$. Hence, we have

$$\begin{aligned}\mathbf{g}_1 + \jmath\mathbf{g}_2 &= (\mathbf{C}\mathbf{q}_1 - \mathbf{D}\mathbf{q}_2) + \jmath\,(\mathbf{D}\mathbf{q}_1 + \mathbf{C}\mathbf{q}_2)\\ &= (\mathbf{C} + \jmath\mathbf{D})\,(\mathbf{q}_1 + \jmath\mathbf{q}_2)\\ &= (\mathbf{A} + \jmath\mathbf{B})^{-1}(\mathbf{q}_1 + \jmath\mathbf{q}_2)\end{aligned} \tag{A.3}$$

Appendix B.

In this appendix, we provide an explanation of $\tilde{\mathbf{E}}^{(j,i)^{-1}}$. For simplicity, we assume that the DF data symbols are all correct, i.e., $\hat{\mathbf{X}}[k] = \mathbf{X}[k]$, and neglect noise terms. Therefore, the LS estimate in (20) for channel $H^{(j,i)}[k]$ given in (1) becomes

$$C\,[k]\,H^{(j,i)}\,[k] = C\,[k]\sum_{l=1}^{G}\breve{\mu}_l^{(j,i)}e^{-\jmath 2\pi(k-1)(l-1)/K} \tag{B.1}$$

for $k \in \mathbf{\Theta}$, where $C[k] = \sum_{t=1}^{N_L}|X^{(i)}[t,k]|^2$ and $\breve{\mu}_l^{(j,i)}$ is the complex gain of the lth path. Taking the IDFT of (B.1), we get the estimate for the $l'th$ channel path gain as follows

$$\hat{\eta}_{l'}^{(j,i)} = \sum_{l=1}^{G}\breve{\mu}_l^{(j,i)}\sum_{k\in\mathbf{\Theta}}C\,[k]\,e^{\jmath 2\pi(k-1)\left(l'-l\right)/K} \tag{B.2}$$

where $1 \le l' \le G$. By rewriting (B.2) in a vector form, we have

$$\hat{\boldsymbol{\eta}}^{(j,i)} = \frac{1}{2}\tilde{\mathbf{E}}^{(j,i)}\breve{\boldsymbol{\mu}}^{(j,i)} \tag{B.3}$$

where $\tilde{\mathbf{E}}^{(j,i)}$ is defined as in (21), $\hat{\boldsymbol{\eta}}^{(j,i)} = [\hat{\eta}_1^{(j,i)}, \ldots, \hat{\eta}_G^{(j,i)}]^T$, and $\breve{\boldsymbol{\mu}}^{(j,i)} = [\breve{\mu}_1^{(j,i)}, \ldots, \breve{\mu}_G^{(j,i)}]^T$. As can be seen in (B.2) and (B.3), $\breve{\mu}_l^{(j,i)}$ from other paths causes interference in $\hat{\eta}_{l'}^{(j,i)}$ due to the effect of aliasing, and $\tilde{\mathbf{E}}^{(j,i)^{-1}}$ acts as a path decorrelator to mitigate this effect.

References

[1] S. Werner, M. Enescu, and V. Koivunen, "Low-complexity time-domain channel estimators for mobile wireless OFDM systems," in *Proc. Signal Processing Systems Design and Implementation Workshop*, Nov. 2005, pp. 245-250.

[2] J.-H. Park, M.-K. Oh, and D.-J. Park, "New channel estimation exploiting reliable decision-feedback symbols for OFDM Systems, " in *Proc. Int. Conf. on Commun.*, June 2006, pp. 3046-3051.

[3] L. Deneire, P. Vandenameele, L. van der Perre, B. Gyselinckx, and M. Engels, "A low-complexity ML channel estimator for OFDM," *IEEE Trans. Commun.*, vol. 51, no. 2, pp. 135-140, Feb. 2003.

[4] M. Morelli and U. Mengali, "A comparison of pilot-aided channel estimation methods for OFDM systems," *IEEE Trans. Signal Processing*, vol. 49, no. 12, pp. 3065-3073, Dec. 2001.

[5] O. Edfors, M. Sandell, J. J. van de Beek, S. K. Wilson, and P. O. Borjesson, "Analysis of DFT-based channel estimators for OFDM," *Wireless Personal Commun.*, vol. 12, no. 1, pp. 55-70, Jan. 2000.

[6] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1451-1458, Oct. 1998.

[7] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, no. 2, pp. 744-765, Mar. 1998.

[8] Jasbir S. Arora, *Introduction to Optimum Design. Elsevier*, 2004.

[9] J. Laiho, A. Wacker, and T. Novosad, *Radio Network Planning and Optimisation for UMTS*. New York: Wiley, 2002.

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

[11] M.-L. Ku and C.-C. Huang, "A refined channel estimation method for STBC/OFDM systems in high-mobility wireless channels," *IEEE Trans. Wireless Commun.*, accepted.

Meng-Lin Ku was born in Taoyuan, Taiwan, R.O.C. He received the B.S. and M.S. degrees in communication engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2002 and 2003, respectively, where he is currently working towards the Ph.D. degree in the Department of Communication Engineering. His research interests are in the area of digital receiver design and optimization for wireless communication systems.

Chia-Chi Huang was born in Taiwan, R.O.C. He received the B.S. degree in electrical engineering from National Taiwan University in 1977 and the M.S. and ph.D. degrees in electrical engineering from the University of California, Berkeley, in 1980 and 1984, respectively.

From 1984 to 1988, he was an RF and Communication System Engineer with the Corporate Research and Development Center, General Electric Co., Schenectady, NY, where he worked on mobile radio communication system design. From 1989 to 1992, he was with the IBM T.J. Watson Research Center, Yorktown Heights, NY, as a Research Staff Member, working on indoor radio communication system design. Since 1992, he has been with the Department of Communication Engineering, National Chiao Tung University, Hsinchu, Taiwan, currently as a Professor.