Spectrally Efficient Time-Frequency Training OFDM for Mobile Large-Scale MIMO Systems

Linglong Dai, Zhaocheng Wang, and Zhixing Yang

***Abstract*—Large-scale orthogonal frequency division multiplexing (OFDM) multiple-input multiple-output (MIMO) is a promising candidate to achieve the spectral efficiency up to several tens of bps/Hz for future wireless communications. One key challenge to realize practical large-scale OFDM MIMO systems is high-dimensional channel estimation in mobile multipath channels. In this paper, we propose the time-frequency training OFDM (TFT-OFDM) transmission scheme for large-scale MIMO systems, where each TFT-OFDM symbol without cyclic prefix adopts the time-domain training sequence (TS) and the frequency-domain orthogonal grouped pilots as the time-frequency training information. At the receiver, the corresponding time-frequency joint channel estimation method is proposed to accurately track the channel variation, whereby the received time-domain TS is used for path delays estimation without interference cancellation, while the path gains are acquired by the frequency-domain pilots. The channel property that path delays vary much slower than path gains is further exploited to improve the estimation performance, and the sparse nature of wireless channel is utilized to acquire the path gains by very few pilots. We also derive the theoretical Cramér-Rao lower bound (CRLB) of the proposed channel estimator. Compared with conventional large-scale OFDM MIMO systems, the proposed TFT-OFDM MIMO scheme achieves higher spectral efficiency as well as the coded bit error rate performance close to the ergodic channel capacity in mobile environments.**

***Index Terms*—large-scale MIMO, OFDM, spectral efficiency, time-frequency training (TFT), time-frequency joint channel estimation.**

I. Introduction

ORTHOGONAL frequency division multiplexing (OFDM) and multiple-input multiple-output (MIMO) are widely recognized as two fundamental physical layer technologies for future wireless communications due to the outstanding capability to combat multipath fading and high spectral efficiency [1], [2]. Currently, most of the literature as well as wireless standards (e.g., IEEE 802.11n [3], IEEE 802.16m [4], 3GPP long term evolution (LTE) [5], etc.) mainly address a small number of transmit antennas (e.g., 2, or 4), whereby the spectral efficiency of about 10 bps/Hz or less can be achieved [1]–[5]. However, large-scale MIMO systems with tens of antennas could achieve the attractive spectral efficiency up to several tens of bps/Hz [6]–[9]. For example, NTT DoCoMo has demonstrated the field experiment of a large-scale MIMO system equipped with 12 $\times$ 12 antennas, which approximately achieves the spectral efficiency of 50 bps/Hz with the transmission rate of 4.92 Gbps over a 100 MHz channel [10]. The evolution of WiFi standard called IEEE 802.11ac is now considering the 16 $\times$ 16 MIMO configuration [11]. In addition, the measurement of 64 $\times$ 64 MIMO channel has also been reported in [12]. The key challenges for realizing large-scale MIMO systems includes proper antenna placement to ensure independent channels, low-complexity signal detection algorithms for practical implementation, channel estimation of the high-dimensional MIMO channel matrix, etc. [6], [8]. This paper will focus on channel estimation for large-scale OFDM MIMO systems.

Basically, there are two categories of channel estimation schemes for OFDM MIMO systems: frequency-domain estimation and time-domain estimation. Normally, frequency-domain estimation can be easily achieved by orthogonal pilots, which converts channel estimation in MIMO systems to that in single-input single-output (SISO) systems [13]. However, the required number of pilots dramatically increases when the number of transmit antennas becomes large. To maintain the pilot overhead within a certain level in OFDM MIMO systems, it is common to reduce the pilot density to some extent, but poor channel estimation performance would be caused, especially in large-scale MIMO systems. Alternatively, time-domain channel estimation based on preamble could provide more reliable estimation over slow fading channels, since all subcarriers in the preamble can be used for channel estimation. However, for large-scale OFDM MIMO systems, the overhead due to preamble is also high [10]. More importantly, unless slow channel fading is assumed, the preamble should be frequently inserted to track the channel variation in mobile environments [14], [15], since high-quality services are expected not only at home/office, but also in mobile cars, buses, etc. Furthermore, flat fading channels are normally assumed for large-scale MIMO systems [8], [16], but actual wireless channels are usually characterized by severe frequency-selective fading due to large delay spread, especially in urban/outdoor scenarios [17].

To solve the channeling problem of high-dimensional channel estimation for large-scale OFDM MIMO systems over frequency-selective mobile channels, we propose in this paper the spectrally efficient time-frequency training OFDM (TFT-OFDM) transmission scheme for large-scale MIMO systems.

Manuscript received 3 February 2012; revised 15 June 2012. This work was supported by National Key Basic Research Program of China (Grant No. 2013CB329203), National Natural Science Foundation of China (Grant Nos. 61271266, 61201185, 60902003), China Postdoctoral Science Special Foundation (Grant No. 2012T50093), Science and Technology Foundation for Beijing Outstanding Doctoral Dissertation Supervisor (Grant No. 20121000303), and Tsinghua University-KU Leuven Bilateral Scientific Cooperation Foundation (Grant No. BIL11/21T).

The authors are with Department of Electronic Engineering as well as Tsinghua National Laboratory for Information Science and Technology (TNList), Tsinghua University, Beijing 100084, P. R. China (e-mails: {daill, zcwang, yangzhx}@tsinghua.edu.cn).

Digital Object Identifier 10.1109/JSAC.2013.130213.

Specifically, the contributions of this paper are summarized as follows:

1) Without cyclic prefix (CP) in standard OFDM (which is explicitly referred as "CP-OFDM" in the sequel), we propose a novel OFDM transmission scheme called TFT-OFDM, which jointly adopts the one-sample shifted time-domain training sequence (TS) and the frequency-domain orthogonal grouped pilots as the time-frequency training information for every OFDM data block. The corresponding time-frequency joint channel estimation scheme for TFT-OFDM MIMO is also proposed, whereby the received time-domain TS is directly used for path delays estimation without interference cancellation, while the path gains are acquired by the frequency-domain grouped pilots;
2) We further exploit the channel property that path delays vary much slower than path gains to improve the path delay estimation accuracy by averaging dozens of received TSs, and employ the sparse nature of wireless channels to acquire the path gains by only a very small amount of pilots. Therefore, the proposed scheme achieves higher spectral efficiency than frequency-domain based channel estimations where plenty of pilots are required to estimate both path delays and path gains;
3) We also propose the TFT-OFDM based transmission frame composed of one preamble and the subsequent TFT-OFDM subframes, whereby the preamble is used to provide an initial reliable channel estimate. Unlike the preamble-based time-domain channel estimation schemes rely heavily on the preamble, channel tracking of the proposed TFT-OFDM MIMO scheme is achieved by time-frequency joint channel estimation realized in every TFT-OFDM subframe. Thus, compared with the time-domain preamble based solutions, the preamble in the proposed scheme can be used much less frequently in mobile environments, leading to the higher spectral efficiency of the TFT-OFDM MIMO scheme;
4) Moreover, the Cramér-Rao lower bound (CRLB) of the proposed channel estimator is derived in this paper, which is approached by simulation results. We also show that TFT-OFDM enjoys the near-capacity performance much better than conventional schemes over doubly-selective channels as well as the robustness to mobile channels.

The rest of this paper is organized as follows. The system model of the proposed TFT-OFDM MIMO scheme is presented in Section II. The preamble-based channel estimators in both the time and frequency domains are discussed in Section III. Section IV addresses the channel tracking and data detection for TFT-OFDM MIMO systems. The performance analysis of the proposed scheme is provided in Section V before simulation results are presented in Section VI. Finally, conclusions are drawn in Section VII.

Notation: The boldface letters are used to denote matrices and vectors. The upper- and lower-case characters are used to represent quantities in the frequency and time domains, respectively. $\mathbf{F}_N$ is the $N \times N$ normalized discrete Fourier transform (DFT) matrix with the $(n+1, k+1)$th entry $\exp(-j2\pi nk/N)/\sqrt{N}$. $\mathbf{I}_N$ is the $N \times N$ identity matrix, and $\mathbf{0}_{M\times N}$ is the $M \times N$ zero matrix. $\otimes$ means the circular correlation. $(\cdot)^*$, $(\cdot)^{\mathrm{T}}$, $(\cdot)^H$, $(\cdot)^{-1}$, $(\cdot)^\dagger$ and $|\cdot|$ denote the complex conjugate, transpose, conjugate transpose, matrix inversion, Moore-Penrose matrix inversion and absolute operations, respectively. $\mathrm{Tr}\{\cdot\}$, $\mathrm{E}\{\cdot\}$, and $\det\{\cdot\}$ stand respectively for trace, expectation, and determinant operators. $\widehat{x}$ means the estimate of x. Finally, $\mathrm{diag}\{\mathbf{u}\}$ is a diagonal matrix with $\mathbf{u}$ at its main diagonal.

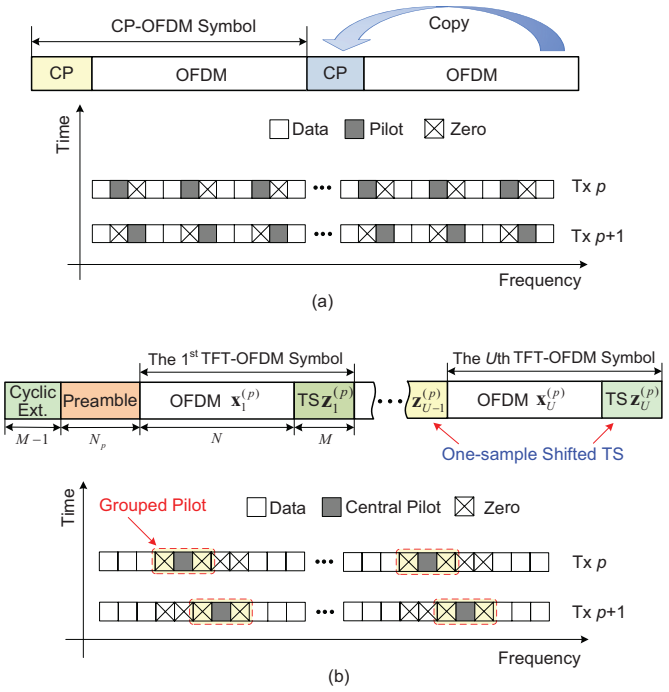

Fig. 1. Time-frequency signal structure comparison: a) CP-OFDM MIMO with frequency-domain training information (pilots) only; b) The proposed TFT-OFDM MIMO with both time- and frequency-domain training information for every OFDM data block.

II. System Model

In this section, the time-frequency signal structure of the proposed TFT-OFDM MIMO scheme is described at first, and then the system model is presented.

A. Time-Frequency Signal Structure of TFT-OFDM

Fig. 1 compares the time-frequency signal structure of CP-OFDM MIMO and the proposed TFT-OFDM MIMO. As shown in Fig. 1 (b), the TFT-OFDM signals are transmitted frame by frame, whereby each frame is composed of one preamble with its cyclic extension and the following U TFT-OFDM symbols (subframes). We assume N_t transmit antennas and N_r receive antennas in MIMO systems.

In the time domain, unlike CP-OFDM where the CP is utilized as the guard interval, the ith TFT-OFDM symbol $\mathbf{s}_i^{(p)}$ $(1 \le i \le U)$ for the pth transmit antenna $(1 \le p \le N_t)$ is composed of the length-N OFDM symbol $\mathbf{x}_i^{(p)} = [x_{i,0}^{(p)}, x_{i,1}^{(p)}, \cdots, x_{i,N-1}^{(p)}]^T$ and the followed length-M

TS $\mathbf{z}_i^{(p)} = [z_{i,0}^{(p)}, z_{i,1}^{(p)}, \cdots, z_{i,M-1}^{(p)}]^T$ as below

$$\mathbf{s}_i^{(p)} = \left[\begin{array}{c} \mathbf{x}_i^{(p)} \\ \mathbf{z}_i^{(p)} \end{array} \right]_{(N+M)\times 1} = \left[\begin{array}{c} \mathbf{F}_N^H \mathbf{X}_i^{(p)} \\ \mathbf{z}_i^{(p)} \end{array} \right]_{(N+M)\times 1}, \quad (1)$$

where $\mathbf{X}_i^{(p)} = \mathbf{F}_N \mathbf{x}_i^{(p)}$. Since it has been proved in [18] that constant TS within the transmission frame is not optimal for channel tracking, the TS $\mathbf{z}_i^{(p)}$ of the ith TFT-OFDM symbol will be generated by cyclically shifting the basic TS $\mathbf{z}^{(p)} = [z_0^{(p)}, z_1^{(p)}, \cdots, z_{M-1}^{(p)}]^T$ by i samples to the left according to

$$\mathbf{z}_i^{(p)} = \left[\begin{array}{cc} \mathbf{0}_{(M-i)\times i} & \mathbf{I}_{M-i} \\ \mathbf{I}_i & \mathbf{0}_{i\times(M-i)} \end{array} \right] \mathbf{z}^{(p)}, \quad (2)$$

where $\mathbf{z}^{(p)}$ for the pth transmit antenna is the Zadoff-Chu sequence (also known as generalized Chirp-like (GCL) sequence) [19] defined by

$$z_m^{(p)} = \exp\left(j\frac{M-1}{M}\pi m^2 r_p\right), 0 \le m \le M-1, \quad (3)$$

where r_p is relatively prime to M. The Zadoff-Chu sequences have ideal autocorrelation and the optimal crosscorrelation equals to the theoretical Sarwate bound [20]. In addition, the Zadoff-Chu sequences with constant envelope both in the time and frequency domains achieve the lowest peak-to-average power ration (PAPR) of 0 dB. Moreover, according to (2), we have

$$\mathbf{z}_{i+1}^{(p)} = [z_{i,1}^{(p)}, z_{i,2}^{(p)}, \cdots, z_{i,M-1}^{(p)}, z_{i,0}^{(p)}]^T, \quad (4)$$

which indicates that the TSs $\mathbf{z}_i^{(p)}$ and $\mathbf{z}_{i+1}^{(p)}$ for two adjacent TFT-OFDM symbols are cyclically shifted to the left by one sample, i.e., the first $M-1$ samples of $\mathbf{z}_{i+1}^{(p)}$ is identical with the last $M-1$ samples of $\mathbf{z}_i^{(p)}$. Thus, the one-sample shifted non-constant TSs $\{\mathbf{z}_i^{(p)}\}_{i=1}^U$ in TFT-OFDM could also preserve the cyclic property of the transmitted data stream when constant TSs are used instead. This property is useful for low-complexity equalization [21] as well as accurate timing/frequency synchronization [22].

In the frequency domain, TFT-OFDM MIMO systems adopts G orthogonal pilot groups randomly scattered within the signal bandwidth, where each pilot group has only one non-zero central pilot in the middle surrounded by d zero pilots on the left and right sides (in Fig. 1 (b), $d = 1$ is used as an example). Although frequency-domain pilots are common in OFDM systems, the proposed pilot pattern has the following four distinct features: 1) In most OFDM systems, the pilots are regularly placed in the frequency domain, e.g., the equally-spaced comb-type pilots are used to achieve the optimal channel estimation performance [23], while the proposed pilots are randomly inserted instead; 2) The $2d$ zero pilots are used to alleviate the potential inter-carrier-interference (ICI) imposed on the central pilot over fast fading channels. Since ICI is dominantly caused by the neighboring subcarriers [24], $d = 1$ could be used for TFT-OFDM even the channel is varying fast. Note that for large-scale MIMO systems typically used in mobile channels with not very high speed, $d = 0$ can be adopted; 3) The subcarrier index set of the central pilots can be denoted by

$$\mathbf{G}^{(p)} = [g_0 + O_p, g_1 + O_p, \cdots, g_{G-1} + O_p]^T, \quad (5)$$

where $O_p = (d+1)(p-1)$ is the subcarrier offset for the pth transmit antenna. To ensure the orthogonality of the grouped pilots associated with all N_t transmit antennas, $(N_t-1)(2d+1)$ extra zero pilots are usually required for each grouped pilot having $2d+1$ pilots. However, due to the existence of the $2d$ zero pilots around the central pilot, the grouped pilots for adjacent two transmit antennas could overlap d zero pilots, e.g., only $(N_t-1)(d+1)$ instead of $(N_t-1)(2d+1)$ extra zero pilots will be required; 4) The $2d$ zero pilots naturally permit pilot power boosting technique [1] because the power originally dedicated to them can be used by the central pilot instead, i.e., the central pilots could be $X_{i,k}^{(p)} = \sqrt{2d+1}$ for $k \in \mathbf{G}^{(p)}$ if the average power of the data subcarriers is $\mathrm{E}\{|X_{i,k}^{(p)}|^2\} = 1$ for $k \notin \mathbf{G}^{(p)}$. However, without loss of generality, the pilot power is assumed to be the same as the useful data, i.e., $X_{i,k}^{(p)} = 1$ for $k \in \mathbf{G}^{(p)}$ in this paper.

The length-N_p basic preamble $\mathbf{c} = \mathbf{F}_{N_p}^H \mathbf{C}$ is also a Zadoff-Chu sequence defined by (3), where N_p (instead of M) and $r_p = 1$ are used. Based on basic preamble $\mathbf{c}$, the preamble for the pth transmit antenna $\mathbf{c}^{(p)} = [c_0^{(p)}, c_1^{(p)}, \cdots, c_{N_p-1}^{(p)}]^T$ is then generated by

$$\mathbf{c}^{(p)} = \mathbf{F}_{N_p}^H \mathbf{C}^{(p)} = \mathbf{F}_{N_p}^H \mathrm{diag}\{\mathbf{C}\}\, \mathbf{W}^{(p)}, \quad (6)$$

where $\mathbf{W}^{(p)} = [0, e^{-j\frac{2\pi}{N_t}(p-1)}, \cdots, e^{-j\frac{2\pi}{N_t}(N_p-1)(p-1)}]^T$, $\mathbf{C}^{(p)} = \mathrm{diag}\{\mathbf{C}\}\, \mathbf{W}^{(p)}$ denotes the FT of $\mathbf{c}^{(p)}$, and $N_p = N_t M$ is assumed. Finally, the cyclic extension $\widetilde{\mathbf{c}}^{(p)}$ of length $M-1$ is generated by

$$\widetilde{\mathbf{c}}^{(p)} = \left[\begin{array}{cc} \mathbf{0}_{(M-1)\times 1} & \mathbf{I}_{M-1} \end{array} \right] \mathbf{c}^{(p)}. \quad (7)$$

■

One remark to make is that, although the preamble insertion is common in OFDM MIMO systems [1], [8], [14], [15] where the preamble-based channel estimation will be used to detect all the following OFDM symbols over slowly time-varying channels[1], it will be shown later in Section V that the preamble in the proposed TFT-OFDM MIMO scheme is used only to acquire the initial channel estimation, while the channel tracking for the subsequent TFT-OFDM symbols are realized by exploiting the time-frequency training information in every TFT-OFDM symbol. Therefore, the subframe number U in the transmission frame could still be large even when the channel is varying fast.

B. System Model of TFT-OFDM MIMO

In MIMO systems, for a certain receive antenna[2], the channel impulse response (CIR) $\mathbf{h}_i^{(p)}$ associated with the pth transmit antenna during the ith TFT-OFDM symbol can be denoted by

$$\mathbf{h}_i^{(p)} = [h_{i,0}^{(p)}, h_{i,1}^{(p)}, \cdots, h_{i,L-1}^{(p)}]^T, \quad (8)$$

[1]Alternatively, the time-domain interpolation between separated preambles can be used to improve the channel tracking capability, but the preamble should be frequently inserted to ensure the reliable performance over fast fading channels.

[2]As the time-frequency training information in every TFT-OFDM symbol will be exploited by every receive antenna adopting the identical processing, the receive antenna index will be omitted from now on.

where $h_{i,l}^{(p)}$ is the path gain of the lth path with the path delay $\tau_l^{(p)}$, L denotes the maximum channel spread, and $L = M$ is assumed to avoid the interference between two neighboring OFDM data blocks, so we have $N_p = N_t M = N_t L$. Note that although the delay spread L maybe large in frequency-selective channels, the number of most significant taps (or resolvable paths) Q is usually much smaller than the channel length L, i.e., $Q \ll L$, because of the sparse nature of wireless channels, especially for wideband communications [25], [26]. For example, the ITU Vehicular B channel [27] with the maximum delay spread of 20 μs, which is equivalent to $L = 200$ samples at the system sampling rate of 10 MHz, has only $Q = 6$ resolvable paths. Furthermore, it has been proved that the path delays $\{\tau_l^{(p)}\}_{l=0}^{L-1}$ vary much slower than path gains $\{h_{i,l}^{(p)}\}_{l=0}^{L-1}$ (including phases and amplitudes) [28], [29], which is caused by the fact that the duration for the delay of a path to change by one tap is inversely proportional to the signal bandwidth B, while the coherence time of the path gains is inversely proportional to the carrier frequency f_c. Since $B \ll f_c$ for almost all of practical wireless systems, path delays would change much slower than path gains[3].

At the receiver, the signals coming from different transmit antennas will mix together, and the received OFDM data block $\mathbf{y}_i = [y_{i,0}, y_{i,1}, \cdots, y_{i,N-1}]^T$ after cyclicity reconstruction [21], [32] is

$$\mathbf{y}_i = \sum_{p=1}^{N_t} \mathbf{x}_i^{(p)} \otimes \mathbf{h}_i^{(p)} + \mathbf{w}_0, \tag{9}$$

where $\mathbf{w}_i = [w_{i,0}, w_{i,1}, \cdots, w_{i,N-1}]^T$ denotes the additive white Gaussian noise (AWGN) vector with zero mean and the variance of $\sigma^2 \mathbf{I}_N$.

Applying DFT to $\mathbf{y}_i$ above, the received signal $Y_{i,k}$ on the kth subcarrier could be presented by

$$Y_{i,k} = \sum_{p=1}^{N_t} X_{i,k}^{(p)} H_{i,k}^{(p)} + W_{i,k}, \quad 0 \le k \le N-1, \tag{10}$$

where $\mathbf{H}_i^{(p)} = [H_{i,0}^{(p)}, H_{i,1}^{(p)}, \cdots, H_{i,N-1}^{(p)}]^T$ is the channel frequency response (CFR) of $\mathbf{h}_i^{(p)}$, and we have

$$\mathbf{H}_i^{(p)} = \sqrt{N} \mathbf{F}_{N,L} \mathbf{h}_i^{(p)}, \tag{11}$$

where $\mathbf{F}_{N,L}$ of size $N \times L$ denotes the first L columns of the DFT matrix $\mathbf{F}_N$. In addition, we use $\mathbf{H}_0^{(p)}$ and $\mathbf{h}_0^{(p)}$ to denote the CFR and CIR during the preamble, respectively.

III. Preamble-Based Channel Estimation

Based on the preamble of the TFT-OFDM transmission frame, the initial channel estimation can be achieved either in the time or frequency domain. Their equivalence will be also proved in this section.

[3]Taking a typical high-definition television (HDTV) wireless system with $B = 8$ MHz and $f_c = 770$ MHz for example [30], the path delays remain constant over hundreds of OFDM symbols, although the path gains may vary after several OFDM symbols. Another example is that, all the channel models provides by numerous communications standards employ the constant path delays [27], [31].

A. Time-Domain Channel Estimation

The received preamble $\mathbf{d}_0 = [d_{0,0}, d_{0,1}, \cdots, d_{0,N_p-1}]^T$ in the time domain at the receive antenna is immune from the inter-block-interference (IBI) due to the protection of the cyclic extension, so $\mathbf{d}_0$ can be expressed by

$$\mathbf{d}_0 = \sum_{p=1}^{N_t} \mathbf{c}^{(p)} \otimes \mathbf{h}_0^{(p)} + \mathbf{v}_0 = \sum_{p=1}^{N_t} \mathbf{c}_0^{(p)} \mathbf{h}_0^{(p)} + \mathbf{v}_0 = \mathbf{c}_0 \mathbf{h}_0 + \mathbf{v}_0, \tag{12}$$

where $\mathbf{c}_0^{(p)}$ is the $N_p \times L$ circulant matrix with the first column being the preamble $\mathbf{c}^{(p)}$, $\mathbf{c}_0 = \left[\mathbf{c}_0^{(1)}, \mathbf{c}_0^{(2)}, \cdots, \mathbf{c}_0^{(N_t)}\right]$ denotes the $N_p \times N_t L$ time-domain training matrix based on $\{\mathbf{c}^{(p)}\}_{p=1}^{N_t}$, $\mathbf{h}_0 = \left[\left(\mathbf{h}_0^{(1)}\right)^T, \left(\mathbf{h}_0^{(2)}\right)^T, \cdots, \left(\mathbf{h}_0^{(N_t)}\right)^T\right]^T$ presents the $N_t L \times 1$ equivalent "total" CIR for all N_t transmit antennas, and $\mathbf{v}_0 = [v_{0,1}, v_{0,1}, \cdots, v_{0,N_p-1}]^T$ stands for the channel's AWGN vector with each element having zero mean and the variance of σ^2.

In (12), there are $N_t L$ unknown parameters in $\mathbf{h}_0$ and N_p observations in $\mathbf{d}_0$. If $N_p \ge N_t L$, the time-domain channel estimate $\widehat{\mathbf{h}}_0$ can be obtained by [33]

$$\widehat{\mathbf{h}}_0 = \mathbf{c}_0^{\dagger} \mathbf{d}_0 = \left(\mathbf{c}_0^H \mathbf{c}_0\right)^{-1} \mathbf{c}_0^H \mathbf{d}_0. \tag{13}$$

We have $\mathrm{E}\left\{\widehat{\mathbf{h}}_0 - \mathbf{h}_0\right\} = \mathrm{E}\left\{\left(\mathbf{c}_0^H \mathbf{c}_0\right)^{-1} \mathbf{c}_0^H \mathbf{v}_0\right\} = \mathbf{0}_{N_t L \times 1}$ due to every element of $\mathbf{v}_0$ has zero mean, so the mean square error (MSE) of the unbiased channel estimator (13) is

$$\begin{aligned} \mathrm{MSE} &= \mathrm{E}\left\{\left(\widehat{\mathbf{h}}_0 - \mathbf{h}_0\right)^H \left(\widehat{\mathbf{h}}_0 - \mathbf{h}_0\right)\right\} \\ &= \mathrm{tr}\left\{\left(\mathbf{c}_0^H \mathbf{c}_0\right)^{-1} \mathbf{c}_0^H \mathrm{E}\left\{\mathbf{v}_0 \mathbf{v}_0^H\right\} \mathbf{c}_0 \left(\mathbf{c}_0^H \mathbf{c}_0\right)^{-1}\right\} \\ &= \sigma^2 \mathrm{tr}\left\{\left(\mathbf{c}_0^H \mathbf{c}_0\right)^{-1}\right\}. \end{aligned} \tag{14}$$

According to the proof in the Appendix, the minimum MSE can be achieved by the following optimal design criterion

$$\mathbf{c}_0^H \mathbf{c}_0 = N_t L \mathbf{I}_{N_t L}. \tag{15}$$

The corresponding MSE in (14) is then derived as[4]

$$\mathrm{MSE}_{\min} = \sigma^2 \mathrm{tr}\left\{\frac{1}{N_t L} \mathbf{I}_{N_t L}\right\} = \sigma^2. \tag{16}$$

It can be verified that the proposed preamble (6) meets the optimal design criterion (15). Thus, the minimum MSE (16) can be achieved, and the time-domain channel estimator (13) is then simplified by circular correlation [33] as

$$\widehat{\mathbf{h}}_0 = \frac{1}{N_t L} \mathbf{c}_0^H \mathbf{d}_0 = \frac{1}{N_t L} \mathbf{c} \otimes \mathbf{d}_0. \tag{17}$$

B. Frequency-Domain Channel Estimation

The frequency-domain signal model (10) is also valid when N_p-point DFT instead of N-point DFT is used to produce the received preamble $\mathbf{D}_0 = [D_{0,0}, D_{0,1}, \cdots, D_{0,N_p-1}]^T$ in the frequency domain, i.e.,

$$\mathbf{D}_0 = \mathbf{C}_0 \mathbf{H}_0 + \mathbf{V}_0, \tag{18}$$

[4]As will be addressed in detail in the Appendix, the results here is consistent with those in [13], but we use distinct proof technique in this paper.

where $\mathbf{V}_0 = \mathbf{F}_{N_p}\mathbf{v}_0$ denotes AWGN, $\mathbf{D}_0 = \mathbf{F}_{N_p}\mathbf{d}_0$ presents the N_p-point DFT of the time-domain received preamble $\mathbf{d}_0$, $\mathbf{C}_0 = \left[\mathrm{diag}\left\{\mathbf{C}^{(1)}\right\}, \mathrm{diag}\left\{\mathbf{C}^{(2)}\right\}, \cdots, \mathrm{diag}\left\{\mathbf{C}^{(N_t)}\right\}\right]$ denotes the $N_p \times N_t N_p$ frequency-domain training matrix based on $\{\mathbf{C}^{(p)}\}_{p=1}^{N_t}$, and the CFR $\mathbf{H}_0$ during the preamble can be related to the corresponding CIR $\mathbf{h}_0$ by using (11) as $\mathbf{H}_0 = \mathbf{F}_0\mathbf{h}_0$, where

$$\mathbf{F}_0 = \begin{bmatrix} \mathbf{F}_{N_p,L} & \cdots & \mathbf{0}_{N_p\times L} \\ \vdots & \ddots & \vdots \\ \mathbf{0}_{N_p\times L} & \cdots & \mathbf{F}_{N_p,L} \end{bmatrix}_{N_tN_p\times N_tL}. \quad (19)$$

Since there are N_tN_p unknown parameters in $\mathbf{H}_0$ and only N_p observations in $\mathbf{D}_0$, eq. (18) is an underdetermined problem without unique solution. However, this problem can be solved by using the relationship between the CFR $\mathbf{H}_0$ and the CIR $\mathbf{h}_0$ as below

$$\mathbf{D}_0 = \mathbf{C}_0\mathbf{F}_0\mathbf{h}_0 + \mathbf{V}_0 = \mathbf{A}_0\mathbf{h}_0 + \mathbf{V}_0, \quad (20)$$

where $\mathbf{A}_0 = \mathbf{C}_0\mathbf{F}_0$, and the number of unknown parameters is reduced from N_tN_p in (18) to N_tL in (20). If $N_p \geq N_tL$, the channel estimation can be achieved by [13]

$$\widehat{\mathbf{h}}_0 = \mathbf{A}_0^{\dagger}\mathbf{D}_0 = \left((\mathbf{C}_0\mathbf{F}_0)^H\mathbf{C}_0\mathbf{F}_0\right)^{-1}(\mathbf{C}_0\mathbf{F}_0)^H\mathbf{D}_0. \quad (21)$$

Then, the CFR can be obtained by $\widehat{\mathbf{H}}_0 = \mathbf{F}_0\widehat{\mathbf{h}}_0 = \mathbf{F}_0\mathbf{A}_0^{\dagger}\mathbf{D}_0$.

C. Unification of The Time- and Frequency-Domain Channel Estimators

The time-domain channel estimator (13) is based on time-domain signals $\mathbf{d}_0$ and $\mathbf{c}_0$, while the frequency-domain channel estimator (21) depends on the frequency-domain signals $\mathbf{D}_0$ and $\mathbf{C}_0$. Since $\mathbf{D}_0$ ($\mathbf{C}_0$) can be obtained once $\mathbf{d}_0$ ($\mathbf{c}_0$) is known, and vice versa, the time- and frequency-domain channel estimators (13) and (21) can be directly unified by the extracted DFT matrix $\mathbf{F}_0$ denoted by (19).

Regarding to the optimal design criterion, we have in Section III-A derived (15) for the time-domain training matrix $\mathbf{c}_0$, while it has been proved in [13] that the optimal frequency-domain channel estimator (21) is subject to the following optimal design criterion

$$\mathbf{A}_0^H\mathbf{A}_0 = N_tL\mathbf{I}_{N_tL}. \quad (22)$$

Using the well-known shift property of DFT, it can be derived that

$$\mathbf{F}_{N_p}\mathbf{c}_0 = \mathbf{C}_0\mathbf{F}_0 = \mathbf{A}_0. \quad (23)$$

With the help of (23), the unification of the optimal design criteria (15) and (22) for the time- and frequency-domain channel estimators, respectively, can be revealed by

$$\mathbf{c}_0^H\mathbf{c}_0 = \left(\mathbf{F}_{N_p}^H\mathbf{A}_0\right)^H\mathbf{F}_{N_p}^H\mathbf{A}_0 = \mathbf{A}_0^H\left(\mathbf{F}_{N_p}\mathbf{F}_{N_p}^H\right)\mathbf{A}_0 = \mathbf{A}_0^H\mathbf{A}_0. \quad (24)$$

It reads clear from (24) that the different design criteria (15) and (22) are essentially equivalent.

Therefore, the time- and frequency-domain channel estimators as well as their corresponding optimal design criteria can be unified under the same framework.

IV. Channel Tracking and Data Detection

The preamble-based initial channel estimation becomes outdated to detect the subsequent TFT-OFDM symbols over fast time-varying channels. This section addresses the cyclicity reconstruction and the channel tracking for every TFT-OFDM symbols within the transmission frame to achieve reliable data detection.

A. Cyclicity Reconstruction of the OFDM Symbol

In CP-OFDM systems, the cyclicity property of the received OFDM symbol is naturally restored due to the insertion of CP. In the proposed TFT-OFDM scheme, the CP is replaced by the known one-sample shifted TS to improve the spectral efficiency. However, the absence of CP would destroy the cyclicity property of the received OFDM symbol over multipath fading channels [21], [32], and cyclicity reconstruction is required to achieve the low-complexity frequency domain equalization similar to that widely used in CP-OFDM.

Due to the multipath propagation, the received OFDM symbol $\mathbf{y}_i$ of the ith TFT-OFDM symbol can be expressed by

$$\mathbf{y}_i = \sum_{p=1}^{N_t}\left(\mathbf{h}_{i,\mathrm{ISI}}^{(p)}\mathbf{x}_i + \mathbf{h}_{i-1,\mathrm{IBI}}^{(p)}\begin{bmatrix}\mathbf{0}_{(N-M)\times 1}\\ \mathbf{z}_{i-1}^{(p)}\end{bmatrix}_{N\times 1}\right) + \mathbf{w}_i, \quad (25)$$

where $\mathbf{h}_{i,\mathrm{ISI}}^{(p)}$ and $\mathbf{h}_{i,\mathrm{IBI}}^{(p)}$ present the $N \times N$ Toeplitz lower and upper triangular matrix with the first column $[h_{i,0}^{(p)}, h_{i,1}^{(p)}, \cdots, h_{i,L-1}^{(p)}, 0, \cdots, 0]^T$ and the first row $[0, \cdots, 0, h_{i,L-1}^{(p)}, h_{i,L-2}^{(p)}, \cdots, h_{i,1}^{(p)}]^T$, respectively, and $\mathbf{w}_i = \mathbf{F}_N^H\mathbf{W}_i$ denotes the AWGN vector. Using the overlap-add (OLA) [34] algorithm by N_t times, the cyclic prefix reconstruction of $\mathbf{y}_i$ can be achieved by

$$\begin{aligned}\widetilde{\mathbf{y}}_i &= \mathbf{y}_i - \sum_{p=1}^{N_t}\left(\mathbf{h}_{i-1,\mathrm{IBI}}^{(p)}\begin{bmatrix}\mathbf{0}_{(N-M)\times 1}\\ \mathbf{z}_{i-1}^{(p)}\end{bmatrix}_{N\times 1}\right) \\ &\quad + \left(\begin{bmatrix}\mathbf{d}_i\\ \mathbf{0}_{(N-M)\times 1}\end{bmatrix} - \sum_{p=1}^{N_t}\mathbf{h}_{i,\mathrm{ISI}}^{(p)}\begin{bmatrix}\mathbf{z}_i^{(p)}\\ \mathbf{0}_{(N-M)\times 1}\end{bmatrix}\right) \\ &= \sum_{p=1}^{N_t}\mathbf{h}_{i,\mathrm{CIR}}^{(p)}\mathbf{x}_i^{(p)} + \mathbf{w}_i + \begin{bmatrix}\mathbf{v}_i\\ \mathbf{0}_{(N-M)\times 1}\end{bmatrix}.\end{aligned} \quad (26)$$

where $\mathbf{d}_i$ is the time-domain received TS of the ith TFT-OFDM symbol, $\mathbf{h}_{i,\mathrm{CIR}}^{(p)} = \mathbf{h}_{i,\mathrm{ISI}}^{(p)} + \mathbf{h}_{i,\mathrm{IBI}}^{(p)}$ means the circular matrix with the first column $[h_0^{(p)}, h_1^{(p)}, \cdots, h_{L-1}^{(p)}, 0, \cdots, 0]^T$. In the first equation of (26), subtracting the second term on the right means removing the IBI caused by the previous TSs $\{\mathbf{z}_{i-1}^{(p)}\}_{p=1}^{N_t}$, and adding the third term indicates restoring the "tails" (buried in $\mathbf{d}_i$) of $\{\mathbf{x}_i^{(p)}\}_{p=1}^{N_t}$ over multipath channels. Since any circular matrix can be diagonalized by the DFT matrix [35], i.e., $\mathbf{h}_{i,\mathrm{CIR}}^{(p)} = \mathbf{F}_N^H\mathbf{H}_i^{(p)}\mathbf{F}_N$, applying DFT to $\widetilde{\mathbf{y}}_i$ would produce the signal $\mathbf{Y}_i$ in (10) as

$$\begin{aligned}\mathbf{Y}_i &= \mathbf{F}_N\widetilde{\mathbf{y}}_i = \sum_{p=1}^{N_t}\mathbf{F}_N\mathbf{F}_N^H\mathbf{H}_i^{(p)}\mathbf{F}_N\mathbf{x}_i^{(p)} + \mathbf{W}_i + \mathbf{v}_i^{'} \\ &= \sum_{p=1}^{N_t}\mathbf{H}_i^{(p)}\mathbf{X}_i^{(p)} + \mathbf{W}_i + \mathbf{v}_i^{'} = \mathbf{X}_i\mathbf{H}_i + \mathbf{W}_i + \mathbf{v}_i^{'},\end{aligned} \quad (27)$$

where $\mathbf{X}_i = \left[\text{diag}\left\{\mathbf{X}_i^{(1)}\right\}, \text{diag}\left\{\mathbf{X}_i^{(2)}\right\}, \cdots, \text{diag}\left\{\mathbf{X}_i^{(N_t)}\right\}\right]$, $\mathbf{H}_i = \left[\left(\mathbf{H}_i^{(1)}\right)^T, \left(\mathbf{H}_i^{(2)}\right)^T, \cdots, \left(\mathbf{H}_i^{(N_t)}\right)^T\right]^T$, $\mathbf{W}_i = \mathbf{F}_N \mathbf{w}_i$, and $\mathbf{v}_i^{'} = \mathbf{F}_N \left[\mathbf{v}_i^T \ \ \mathbf{0}_{1\times(N-M)}\right]^T$ is the additional AWGN introduced by the cyclicity reconstruction, whereby an extra noise term $\mathbf{v}_i$ buried in $\mathbf{d}_i$ has been involved. However, this would lead to negligible SNR loss of $10\log_{10}\left(\frac{M+N}{N}\right)$ at the receiver since the TS length is usually much smaller than the OFDM data block length, e.g., the equivalent SNR loss is 0.26 dB when $M = N/16$.

Although similar to the cyclicity reconstruction method for SISO time-domain synchronous OFDM (TDS-OFDM) systems in [32], [36], the proposed algorithm (26) differs from those of [32], [36] in two aspects. Firstly, the conventional method in SISO scenarios is extended to be used in MIMO scenarios. Secondly and more importantly, the cyclicity reconstruction (26) is completed by only one step, while [32], [36] requires iterative processing (the iteration times is usually three or more).

In practical cyclicity reconstruction implementation, the actual CIR $\mathbf{h}_{i-1}^{(p)}$ in (26) should be replaced by the estimate $\widehat{\mathbf{h}}_{i-1}^{(p)}$, which has been obtained in the previous $(i-1)$th TFT-OFDM symbol, and $\mathbf{h}_i^{(p)}$ in (26) can be either simply approximated by $\widehat{\mathbf{h}}_{i-1}^{(p)}$ or predicted by the Kalman filter exploiting the temporal correlation nature of the channel as well as the previous channel estimates $\{\widehat{\mathbf{h}}_u^{(p)}\}_{u=0}^{i-1}$ [37]. It is worth noting that the approximation of $\mathbf{h}_i^{(p)}$ for cyclicity reconstruction wound indeed incur some performance degradation due to the imperfect matching, but the performance loss is small due to the following two reasons: 1) As will be shown later, the time-frequency joint channel tracking method could achieve the CIR estimate with high accuracy in every TFT-OFDM symbol; 2) Since large-scale MIMO technique is mainly used in static or slow time-varying channels [6], [8]–[11], the CIR estimates in previous TFT-OFDM symbols could be efficiently exploited to produce a good approximation of the actual CIR in the current TFT-OFDM symbol.

B. Time-Frequency Joint Channel Estimation

In most OFDM MIMO systems, channel estimation is achieved in the time or frequency domain by using the time-domain preamble [8], [10] or the frequency-domain pilots [2], [13], [23], respectively. However, taking advantage of the time-frequency training of TFT-OFDM, channel tracking for TFT-OFDM can be realized by the following time-frequency joint channel estimation composed of two sequential steps: the TS-based *path delay* estimation and the pilot-based *path gain* estimation.

The received TS $\mathbf{d}_i = [d_{i,0}, d_{i,1}, \cdots, d_{i,M-1}]^T$ of the ith TFT-OFDM symbol is given by

$$\mathbf{d}_i = \sum_{p=1}^{N_t} \left(\underline{\mathbf{h}}_{i,\text{ISI}}^{(p)} \mathbf{z}_i^{(p)} + \underline{\mathbf{h}}_{i,\text{IBI}}^{(p)} \mathbf{x}_{i,N-M:N-1}^{(p)}\right) + \mathbf{v}_i, \qquad (28)$$

where $\underline{\mathbf{h}}_{i,\text{ISI}}^{(p)}$ and $\underline{\mathbf{h}}_{i,\text{IBI}}^{(p)}$ denote the $M \times M$ Toeplitz lower and upper triangular matrix with the first column $[h_{i,0}^{(p)}, h_{i,1}^{(p)}, \cdots, h_{i,L-1}^{(p)}, 0, \cdots, 0]^T$ and the first row $[0, \cdots, 0, h_{i,L-1}^{(p)}, h_{i,L-2}^{(p)}, \cdots, h_{i,1}^{(p)}]^T$, respectively, and $\mathbf{x}_{i,N-M:N-1}^{(p)}$ presents the last M elements of $\mathbf{x}_i^{(p)}$. The term $\underline{\mathbf{h}}_{i,\text{IBI}}^{(p)} \mathbf{x}_{i,N-M:N-1}^{(p)}$ in (28) indicate that the received TS $\mathbf{d}_i$ is "contaminated" by the IBI caused by the preceding OFDM data blocks.

Being different from TDS-OFDM where iterative IBI cancellation is exploited to achieve the time-domain complete CIR estimation based on the received TS [36], by exploiting the ideal autocorrelation property of the Zadoff-Chu sequences [19], we directly correlate the local Zadoff-Chu sequence $\mathbf{z}_i^{(p)}$ with the "contaminated" TS $\mathbf{d}_i$ without interference cancellation to just estimate the path delays of the channel as below

$$\widehat{\mathbf{h}}_i^{(p)} = \frac{1}{M} \mathbf{z}_i^{(p)} \otimes \mathbf{d}_i = \mathbf{h}_i^{(p)} + \mathbf{n}_i^{(p)} + \mathbf{v}_i^{(p)}, \quad 1 \le p \le N_t, \ (29)$$

where $\mathbf{v}_i^{(p)} = \frac{1}{M} \mathbf{z}_i^{(p)} \otimes \mathbf{v}_i$, and $\mathbf{n}_i^{(p)}$ denotes the interferences caused by the non-zero crosscorrelation among different Zadoff-Chu sequences as well as the IBIs from previous OFDM symbols. As demonstrated later in Fig. 2, if only one received TS is used, due to the absence of interference cancellation, the path delay information hidden in $\widehat{\mathbf{h}}_i^{(p)}$ maybe not accurate when interferences are severe, although the paths with high gains can be identified well. Based on property of wireless channels that path delay varies much slower than path gains [28], [29], we further propose the averaged path delay estimation to further improve the accuracy, whereby the received TSs within β adjacent TFT-OFDM symbols during which the path delays do not change obviously are averaged as below:

$$\overline{\mathbf{h}}_i^{(p)} = \frac{1}{\beta} \sum_{u=i-\beta+1}^{i} \widehat{\mathbf{h}}_u^{(p)} = \frac{1}{\beta M} \mathbf{z}_u^{(p)} \otimes \left(\sum_{u=i-\beta+1}^{i} \mathbf{d}_u\right). \ (30)$$

Note that the number β could be very large (e.g., $\beta > 10$) due to the channel property mentioned above.

Then, the path gains in $\overline{\mathbf{h}}_i^{(p)}$ are directly discarded since they maybe not accurate due to absence of interference cancellation, and only the path delays of the Q most significant taps of $\overline{\mathbf{h}}_i^{(p)}$ are stored in the path delay set

$$\mathbf{\Gamma}_i^{(p)} = \{\tau_l^{(p)} : \left|\overline{h}_{i,l}^{(p)}\right|^2 \ge T_{th}\}_{l=0}^{L-1}, \quad 1 \le p \le N_t, \qquad (31)$$

where T_{th} is the power threshold could be determined according to [38], and Q is usually much smaller than the channel length L, i.e., $Q \ll L$ [25], [26]. That is to say, the number of unknown parameters in the CIR $\mathbf{h}_i^{(p)}$ is substantially reduced from L to Q ($Q \ll L$) after the path delays $\mathbf{\Gamma}_i^{(p)}$ have been obtained. Notice that the path delay estimation of the most significant taps could be very accurate since β could be very large in (30) due to the fact that path delays vary much slower than path gains [28], and furthermore the IBIs caused by preceding OFDM data blocks may be averaged out due to the randomness nature of the data.

After the path delays $\{\mathbf{\Gamma}_i^{(p)}\}_{p=1}^{N_t}$ has been obtained in (28), the number of unknown parameters of $\mathbf{h}_i{}^{(p)}$ is substantially

reduced from L to Q, thus only a small amount of frequency-domain grouped pilots will be sufficient to estimate the Q channel path gains. ■

Due to the orthogonality of the grouped pilots among different transmit antennas, the received central pilots $Y_{i,k}^{(p)}$ can be expressed by

$$Y_{i,k} = \sum_{p=1}^{N_t} X_{i,k}^{(p)} H_{i,k}^{(p)} + W_{i,k} = H_{i,k}^{(p)} + W_{i,k}, \quad k \in \mathbf{G}^{(p)}, \tag{32}$$

where the signal model (10), the central pilots $X_{i,k}^{(p)} = 1$ ($k \in \mathbf{G}^{(p)}$), and $X_{i,k}^{(u)} = 0$ ($u \neq p, k \in \mathbf{G}^{(p)}$) have been utilized. Eq. (32) can be rewritten in a more compact matrix form as

$$\mathbf{Y}_i^{(p)} = \mathbf{F}_N^{(p)} \mathbf{h}_{i,\Gamma}^{(p)} + \mathbf{W}_i^{(p)}, \quad 1 \leq p \leq N_t, \tag{33}$$

where $\mathbf{Y}_i^{(p)} = [Y_{i,g_0+O_p}, Y_{i,g_1+O_p}, \cdots, Y_{i,g_{G-1}+O_p}]_{G\times 1}^T$, $\mathbf{h}_{i,\Gamma}^{(p)} = [h_{i,\tau_0}^{(p)}, h_{i,\tau_1}^{(p)}, \cdots, h_{i,\tau_{Q-1}}^{(p)}]_{Q\times 1}^T$ denotes the "selected" CIR out of $\mathbf{h}_i^{(p)}$ according to the path delays $\mathbf{\Gamma}_i^{(p)}$ obtained in (31), $\mathbf{F}_N^{(p)}$ presents the $G \times Q$ "extracted" matrix generated by selecting the $\mathbf{G}^{(p)}$ rows and $\mathbf{\Gamma}_i^{(p)}$ columns out of the DFT matrix $\mathbf{F}_N$, and $\mathbf{W}_i^{(p)} = [W_{i,g_0+O_p}^{(p)}, W_{i,g_1+O_p}^{(p)}, \cdots, W_{i,g_{G-1}+O_p}^{(p)}]_{G\times 1}^T$ is the AWGN vector. Note that the relationship (11) has been used to obtain (33).

It is clear from (33) that only Q unknown path gains in $\mathbf{h}_{i,\Gamma}^{(p)}$ have to be estimated by the G observations in $\mathbf{Y}_i^{(p)}$. If $G \geq Q$, the Q (not L, and usually $Q \ll L$) path gains of $\mathbf{h}_i^{(p)}$ can be estimated by the received central pilots $\mathbf{Y}_i^{(p)}$ as

$$\widehat{\mathbf{h}}_{i,\Gamma}^{(p)} = \left(\mathbf{F}_N^{(p)}\right)^{\dagger} \mathbf{Y}_i^{(p)} = \left[\left(\mathbf{F}_N^{(p)}\right)^H \mathbf{F}_N^{(p)}\right]^{-1} \left(\mathbf{F}_N^{(p)}\right)^H \mathbf{Y}_i^{(p)}. \tag{34}$$

Note that the pilot power boosting technique [1], whereby pilots use higher power than normal data, is more suitable for TFT-OFDM than CP-OFDM, since no obvious equivalent signal-to-noise ratio (SNR) degradation will be introduced in TFT-OFDM having much fewer pilots than CP-OFDM.

Combining the path delays $\{\mathbf{\Gamma}_i^{(p)}\}_{p=1}^{N_t}$ obtained in (31) based on the time-domain received TS $\mathbf{d}_i$ and the path gains $\{\widehat{\mathbf{h}}_{i,\Gamma}^{(p)}\}_{p=1}^{N_t}$ acquired in (34) based on the frequency-domain received pilots $\{\mathbf{Y}_i^{(p)}\}_{p=1}^{N_t}$, the complete CIR estimates $\{\widehat{\mathbf{h}}_i^{(p)}\}_{p=1}^{N_t}$ for all the N_t transmit antennas could be achieved by the proposed time-frequency joint channel estimation.

C. Data Detection for TFT-OFDM

After cyclicity reconstruction and the time-frequency joint channel estimation have been completed, the obtained OFDM symbol $\mathbf{Y}_i$ and the complete CIR estimates $\{\widehat{\mathbf{h}}_i^{(p)}\}_{p=1}^{N_t}$ are fed into the MIMO detector to recover the transmitted signals. In contrast to conventional detection algorithms for MIMO systems of small size, e.g., maximum-likelihood (ML) or maximum a posteriori (MAP) detectors whose complexity increases exponentially with the number of transmit antennas [8], we resort to the recently proposed low-complexity detection algorithms specially designed for large-scale MIMO systems [8], [16], [39]–[42]. Among them, the first category algorithms based on the local neighborhood search includes the likelihood ascent search (LAS) [8], [39] and reactive tabu search (RTS) [16], which are aimed to achieve the near-ML performance. The second category algorithms approaching the near-MAP performance are based on message passing, e.g., the belief propagation (BP) algorithm [40], [41] and probabilistic data association (PDA) scheme [8], [42]. They all enjoy the preferred property of "large system behavior", i.e., the detection performance improves when the number of transmit antennas N_t becomes large. Since data detection algorithm is not the focus of this paper, we directly adopt the near-capacity RTS scheme with the low per-subcarrier complexity of $\mathcal{O}(N_t N_r)$. Please refer to [16] for more details about the RTS algorithm. Finally, the detected data are used by the channel decoder to ultimately recover the transmitted signal.

V. Performance Analysis

This section addresses the performance analysis of the proposed scheme, including the spectral efficiency of proposed TFT-OFDM scheme, the CRLB as well as the computational complexity of the time-frequency joint channel estimation method.

A. Spectral Efficiency

Due to the overhead caused by the time-domain guard interval (either CP in standard CP-OFDM or TS in the proposed TFT-OFDM) and the frequency-domain pilots, the spectral efficiency η_0 of the proposed TFT-OFDM MIMO scheme normalized by the ideal case without any overhead [21], [43] can be expressed in the percentage notation as

$$\eta_0 = \frac{U(N-K)}{U(N+M)+N_p+M-1}, \tag{35}$$

where $K = G\left((2d+1)+(N_t-1)(d+1)\right)$ is the number of used pilots in each OFDM data block.

For typical wireless digital television systems, large DFT size, e.g., $N = 4096$, is usually adopted [30]. Since all channel models defined by ITU [27] and all channel models used for digital television system evaluation [31] have no more than six resolvable paths, we could assume $G = Q = 6$ without loss of generality. However, in practical applications, the path number may be large, so we configure $G = 10$ for system design with some margin. As mentioned in Section II-A, even we relax the quasi-static channel over the entire transmission frame to only one subframe, $d = 0$ could be used for large-scale MIMO systems where the mobile channels are not varying very fast[5].

Thus, for large-scale 16×16 MIMO configuration, i.e., $N_t = N_r = 16$, the number of used pilots in TFT-OFDM is $K = 160$, which is only 3.91% of the total subcarrier number $N = 4096$. If the channel is varying slowly, e.g., the grouped pilots in TFT-OFDM transmission frame can be inserted every three TFT-OFDM subframes (the corresponding channel estimate update frequency is still much higher than the

[5]Note that the proposed TFT-OFDM can be also used over fast varying channels, whereby $d = 1$ could be configured to alleviate ICI caused by the rapid channel variation [44].

TABLE I
SPECTRAL EFFICIENCY IN MIMO SYSTEMS.

Guard Interval Length	Frequency-domain pilots based MIMO [45]	Time-domain preamble based MIMO [8]	Proposed TFT-OFDM MIMO
$M = N/16$	75%	80%	88.67%

time-domain preamble based scheme [8]), the pilot occupation ratio will be reduced to 1.30%. On the contrary, the Karhunen-Loeve theorem [24] requires that the number of frequency-domain pilot should be not smaller than the channel length L for frequency-domain channel estimation, and the pilot number increases linearly to be N_tL in CP-OFDM MIMO systems due to the orthogonality requirement. This means that $N_tL = 4096$ pilots occupying 100% of the total subcarriers are required in CP-OFDM MIMO systems when the typical guard interval length $M = N/16$ is applied. No extra subcarrier is available if the number of transmit antennas $N_t > 16$. In practical applications [3], [5], [45], fewer pilots can be used due to interpolation can be used to estimate the CFR at data subcarriers at the cost of performance degradation. However, the pilot occupation ratio should be above a certain threshold to ensure reliable performance for large-scale MIMO systems. One typical threshold recommenced by LTE is that 25% subcarriers are used as pilots when $N_t > 4$ [5]. Therefore, we can conclude that the proposed TFT-OFDM MIMO scheme has much higher spectral efficiency than standard CP-OFDM MIMO systems.

Compared with the time-domain preamble based schemes [8], [10] where the preamble should be frequently inserted (or equivalently, small subframe number will be adopted) in mobile environments, the subframe number U in the TFT-OFDM MIMO transmission frame can still be large due to the good channel tracking capability of the time-frequency joint channel estimation in every TFT-OFDM symbol. For example, 4 out of the total 29 OFDM symbols in a frame are dedicated for channel estimation in the 12×12 MIMO experiment with the mobile speed of 10 km/h [10], while the simulation results in Section VI demonstrate that large $U = 50$ could still ensure near-capacity performance for TFT-OFDM MIMO scheme over 30 km/h time-varying channel, whereby $U = 4$ should be used for the time-domain preamble based iterative channel estimation/data detection system [8]. Although the actual spectral efficiency of practical systems may vary with different configurations, with those data addressed above, Table I clearly indicates that the proposed TFT-OFDM MIMO scheme outperforms its conventional counterparts in spectral efficiency.

Since every receive antenna could use the time-frequency joint channel estimation method to distinguish the channels between different transmit antennas and this specific receive antenna, the proposed TFT-OFDM transmission scheme could be used in both scenarios having small (even a single) or large number of receive antennas, and higher spectral efficiency than conventional solutions could be achieved in both cases.

B. Cramér-Rao Lower Bound

The CRLB bound is the theoretical bound to evaluate the performance of practical estimation methods [33]. Since the AWGN vector $\mathbf{W}_i^{(p)}$ in (33) is subject to the distribution of $\mathcal{CN}\left(0, \sigma^2\mathbf{I}_G\right)$, the conditional probability density function (PDF) of $\mathbf{Y}_i^{(p)}$ with the given $\mathbf{h}_{i,\Gamma}^{(p)}$ is

$$p_{\mathbf{Y}_i^{(p)}\left|\mathbf{h}_{i,\Gamma}^{(p)}\right.}\left(\mathbf{Y}_i^{(p)}; \mathbf{h}_{i,\Gamma}^{(p)}\right) = \frac{1}{\left(2\pi\sigma^2\right)^{G/2}} \exp\left\{-\frac{1}{2\sigma^2}\left\|\mathbf{Y}_i^{(p)} - \mathbf{F}_N^{(p)}\mathbf{h}_{i,\Gamma}^{(p)}\right\|^2\right\}. \tag{36}$$

The Fisher information matrix [33] of (33) can then be derived as

$$\begin{aligned}\left[\mathbf{J}\right]_{m,n} &\triangleq -\mathrm{E}\left\{\frac{\partial^2 \ln\left(p_{\mathbf{Y}_i^{(p)}\left|\mathbf{h}_{i,\Gamma}^{(p)}\right.}\left(\mathbf{Y}_i^{(p)}; \mathbf{h}_{i,\Gamma}^{(p)}\right)\right)}{\partial h_{i,\Gamma,m}^{(p)} \partial h_{i,\Gamma,n}^{(p)}}\right\} \\ &= \frac{1}{\sigma^2}\left[\left(\mathbf{F}_N^{(p)}\right)^H \mathbf{F}_N^{(p)}\right]_{m,n},\end{aligned} \tag{37}$$

where $h_{i,\Gamma,m}^{(p)}$ and $h_{i,\Gamma,n}^{(p)}$ denotes the mth and nth entry of $\mathbf{h}_{i,\Gamma}^{(p)}$, respectively. Finally, according to the vector estimation theory [33], the CRLB of the unbiased estimator $\widehat{\mathbf{h}}_{i,\Gamma}^{(p)}$ is

$$\begin{aligned}\mathrm{CRLB} &= \mathrm{E}\left\{\left\|\widehat{\mathbf{h}}_{i,\Gamma}^{(p)} - \mathbf{h}_{i,\Gamma}^{(p)}\right\|^2\right\} \geq \mathrm{Tr}\left\{\mathbf{J}^{-1}\right\} \\ &= \sigma^2 \mathrm{Tr}\left\{\left(\left(\mathbf{F}_N^{(p)}\right)^H \mathbf{F}_N^{(p)}\right)^{-1}\right\}.\end{aligned} \tag{38}$$

Let $\{\lambda_i\}_{i=0}^{Q}$ being the Q eigenvalues of the matrix $\left(\mathbf{F}_N^{(p)}\right)^H \mathbf{F}_N^{(p)}$, then, we have the following result according to the elementary linear algebra [46]

$$\begin{aligned}\mathrm{Tr}\left\{\left(\left(\mathbf{F}_N^{(p)}\right)^H \mathbf{F}_N^{(p)}\right)^{-1}\right\} &= \sum_{i=1}^{Q} \lambda_i^{-1} = Q\left(\sum_{i=1}^{Q} \lambda_i^{-1}/Q\right) \\ &\geq Q\left(Q/\sum_{i=1}^{Q}\lambda_i\right) = \frac{Q^2}{\mathrm{Tr}\left\{\left(\mathbf{F}_N^{(p)}\right)^H \mathbf{F}_N^{(p)}\right\}},\end{aligned} \tag{39}$$

where the equality holds if and only if $\lambda_1 = \lambda_2 = \cdots = \lambda_Q$, which means that the matrix $\mathbf{F}_N^{(p)}$ extracted from the standard DFT matrix $\mathbf{F}_N$ should have orthogonal columns. Obviously, the $Q \times Q$ matrix $\left(\mathbf{F}_N^{(p)}\right)^H \mathbf{F}_N^{(p)}$ has identical diagonals equal to G, i.e., $\mathrm{Tr}\left\{\left(\mathbf{F}_N^{(p)}\right)^H \mathbf{F}_N^{(p)}\right\} = GQ$, so the CRLB of the proposed time-frequency joint channel estimator becomes

$$\mathrm{CRLB} = \mathrm{E}\left\{\left\|\widehat{\mathbf{h}}_{i,\Gamma}^{(p)} - \mathbf{h}_{i,\Gamma}^{(p)}\right\|^2\right\} = \frac{Q\sigma^2}{G}. \tag{40}$$

It is worth noting that if the matrix $\mathbf{F}_N^{(p)}$ does not have orthogonal columns, the MSE of the practical channel estimator will not achieve the CRLB (40). However, due to the random positions of the central pilots in TFT-OFDM, the "extracted" matrix $\mathbf{F}_N^{(p)}$ in the proposed scheme has imperfect but approximate orthogonal columns (or equivalently, $\left(\mathbf{F}_N^{(p)}\right)^H \mathbf{F}_N^{(p)} \approx G\mathbf{I}_Q$), so the CRLB could be asymptotically approached, which will be validated by the simulation results in Section VI. Also note that (40) implies that the increased number of transmit antennas N_t has little impact on the channel tracking performance, but the loss in spectral efficiency has to be paid according to (35).

C. Computational Complexity

The computational complexity of the proposed time-frequency joint channel tracking scheme can be evaluated in terms of how many multiplications are required. The circular correlation of the preamble-based initial channel estimate $\widehat{\mathbf{h}}_0$ in (17) can be efficiently realized by one N_p-point DFT (note that the DFT of the basic preamble $\mathbf{c}$ can be prestored at the receiver) and one N_p-point inverse DFT (IDFT) plus N_p multiplications. In the proposed scheme, $N_p = N_t M$. The cyclicity reconstruction (26) requires N_t times of M-point linear convolution to compute the IBI between the TS and OFDM symbol to obtain $\widetilde{\mathbf{y}}_i$, where each M-point linear convolution can be implemented by one $2M$-point DFT (note that the DFT of the local TSs $\mathbf{z}_i^{(p)}$ and $\mathbf{z}_{i-1}^{(p)}$ are also prestored at the receiver) and one $2M$-point IDFT plus $2M$ multiplications. Then, the N-point DFT is used to produce the frequency-domain OFDM symbol $\mathbf{Y}_i$ in (27). To acquire the TS-based path delay estimates $\{\mathbf{\Gamma}_i^{(p)}\}_{p=1}^{N_t}$ in (31), $\{\overline{\mathbf{h}}_i^{(p)}\}_{p=1}^{N_t}$ in (30) requires N_t times of M-point circular correlation, which can be realized by one M-point DFT of $\mathbf{d}_i$ and N_t times of M-point IDFT since all the DFT of local TSs $\{\mathbf{z}_i^{(p)}\}_{p=1}^{N_t}$ are known to the receiver. The pilot-based path gain estimates $\{\widehat{\mathbf{h}}_{i,\Gamma}^{(p)}\}_{p=1}^{N_t}$ for all N_t transmit antennas in (34) require $N_t(GQ^2+2Q^3)$ multiplications to compute the Moore-Penrose inverse matrix of $\mathbf{F}_N^{(p)}$ and N_tQG multiplications for matrix product. Finally, based on the path delays and path gains for all the N_t transmit antennas, N_t times of N-point DFT is used to convert the CIRs $\{\widehat{\mathbf{h}}_i^{(p)}\}_{p=1}^{N_t}$ to the corresponding CFRs $\{\widehat{\mathbf{H}}_i^{(p)}\}_{p=1}^{N_t}$. Therefore, the overall complexity of proposed channel estimation is $\mathcal{O}\left(N_t N \log_2 N\right) + \mathcal{O}(\alpha N_t)$ where $\alpha = GQ^2 + 2Q^3 + QG$, which is a little higher than the complexity $\mathcal{O}({N_t}^2 N_r) + \mathcal{O}({N_t}^3)$ that of the time-domain based scheme [8] and the complexity $\mathcal{O}\left(N_t M \log_2 M\right) + \mathcal{O}(N_t)$ that of the frequency-domain solution [23] .

VI. Simulation Results

This section investigates the performance of the proposed TFT-OFDM scheme for large-scale MIMO systems in mobile multipath channels. For performance comparison, the proposed TFT-OFDM scheme for large-scale MIMO systems is compared with its conventional counterparts based on CP-OFDM with frequency-domain comb-type pilots [45] and time-domain preamble based iterative channel estimation/data detection scheme in [8], respectively. Note that the spectral efficiency of those three different schemes has been specified in Table I. All those three systems have the same 16×16 MIMO configuration working with the signal bandwidth of 7.56 MHz at the central radio frequency of 770 MHz. The system parameters for TFT-OFDM MIMO scheme are consistent with those specified in Section V-A, e.g., $N = 4096$, $M = 4096/16 = 256$, $U = 10$, $G = 10$, $d = 0$. For all schemes, we adopt the spectrally-efficient non-orthogonal space-time block code (STBC) [47], since it could simultaneously achieve full transmit diversity as well as full rate. As mentioned before, the RTS algorithm [16] is employed for data detection. Since almost all practical OFDM systems use channel coding for reliable performance, we adopt the powerful low-density parity-check (LDPC) code with the block length of 64, 8000 bits and code rate of 2/3 as specified by the standard [30]. The well-known iterative decoding algorithm called belief propagation (BP) [48] is used with the maximum iteration number of 50. The quadrature phase shift keying (QPSK) modulation scheme is simulated. Two typical 6-tap multipath channel models named Brazil D [31] and Vehicular B [27] are used. The first channel has the maximum delay spread of 20 μs, while the later one having the 0 dB echo at the delay of 5.86 μs is deeply frequency-selective characterizing the single frequency network (SFN) environment. The channels associated with different transmit-receive antenna pairs are assumed independent. The mobile velocities of 5 km/h (outdoor walking speed) and 30 km/h (vehicle speed in urban areas) are considered[6]. For the preamble-based transmission scheme with channel estimation errors, the capacity lower bound C has been theoretically proved as [49]

$$C \geq \eta_0 \mathrm{E}\left[\log\det\left(\mathbf{I}_{N_t} + \frac{\gamma^2 M}{N_t(1+\gamma)+\gamma M}\frac{\widehat{\mathbf{H}}\widehat{\mathbf{H}}^H}{N_t\sigma^2_{\widehat{\mathbf{H}}}}\right)\right], \quad (41)$$

where $\sigma^2_{\widehat{\mathbf{H}}} = 1/N_t N_r \mathrm{E}\left[\mathrm{tr}(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^H)\right]$, and γ is the average received SNR. Note that (41) is different from the ideal capacity bound under perfect channel state information (CSI) at the receiver [39], and the minimum SNR for a given capacity C can be obtained from (41).

Firstly, we evaluate the averaged path delay estimation over multipath channels. Fig. 2 and Fig. 3 show the simulation results with the SNR of 5 dB over the Brazil D [31] and Vehicular B channel [27], respectively. The actual channel is also plotted for comparison. We can observe that although the rough channel estimates are not accurate due to the absence of interference cancellation, the path delay information of the actual channel could be preserved well, especially when β becomes large (as discussed in Section II-B, β could be very large, e.g., $\beta = 10$ is adopted in the simulation). It should be pointed out that for channels whose active taps are all of relatively high gains (e.g., the Brazil D channel in Fig. 2), the proposed scheme could accurately identify the path delays even when $\beta = 1$, while for channels having very small active taps (e.g., the Vehicular B channel in Fig. 3, whereby the smallest path has 22 dB lower power than the largest one), large β is required for accurate path delay estimation. Note

[6]The mobile speed of 10 km/h is evaluated in the 12×12 large-scale MIMO experiment in [10].

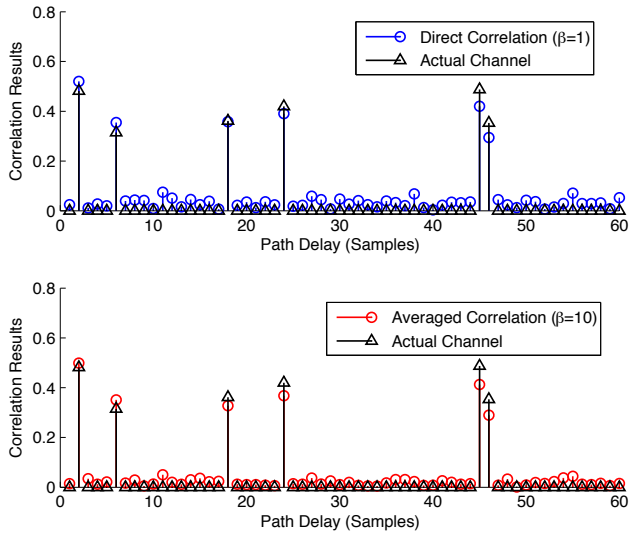

Fig. 2. Averaged path delay estimation over Brazil D channel.

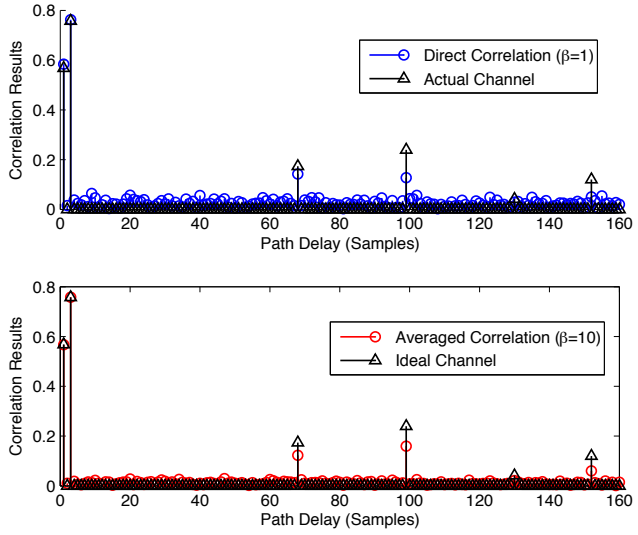

Fig. 3. Averaged path delay estimation over Vehicular B channel.

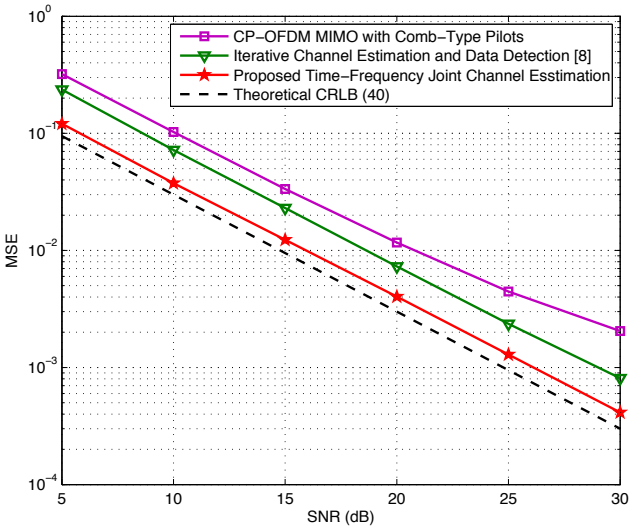

Fig. 4. MSE performance comparison between the proposed time-frequency joint channel estimation method for TFT-OFDM with the conventional schemes.

that if the power of certain path is too small, it can be viewed as noise and will be ignored.

We then investigate in Fig. 4 the MSE performance of the proposed time-frequency joint channel estimation for TFT-OFDM in large-scale MIMO systems over the Brazil D channel with the receiver velocity of 5 km/h. For comparison, we also include the MSE performance of CP-OFDM with comb-type pilots and the time-domain preamble based iterative channel estimation/data detection scheme for large-scale MIMO systems [8]. In addition, the theoretical CRLB derived in (40) is also plotted as the benchmark for comparison. It is clear that TFT-OFDM outperforms CP-OFDM and TDS-OFDM by about 5 dB when the channel estimation MSE 10^{-2} is considered, and performs 3 dB better than [8]. Also we could observe that the proposed channel estimation performs closely to the theoretical CRLB with a small SNR gap, which is caused by the fact that the "extracted" DFT matrix $\mathbf{F}_N^{(p)}$ has imperfect but approximate orthogonal columns.

Fig. 5 compares the LDPC coded bit error rate (BER) performance of TFT-OFDM MIMO system with its counterparts over the Brazil D channel with the receiver velocity of 5 km/h. The BER with perfect ideal channel state information (CSI) is also included as the performance lower bound of practical systems. In addition, as the performance limit, the minimum SNR of 7.1 dB derived to attain the theoretical ergodic capacity (41) for training-based transmission [8] is also plotted as the benchmark for comparison. It is clear that the proposed TFT-OFDM scheme obviously outperforms the conventional schemes dependent on only time or frequency-domain training information. For example, at the coded BER of 10^{-4}, the proposed TFT-OFDM MIMO scheme outperforms CP-OFDM MIMO by the SNR gain of 1.9 dB, and performs 1.4 dB better than the large-scale MIMO scheme with time-domain preamble [8]. We can also find that TFT-OFDM with the proposed time-frequency joint channel estimation is about 1.2 dB away from the ideal CSI case. Moreover, TFT-OFDM performs only about 2.1 dB away from the theoretical bound, which indicates the near-capacity performance of the proposed scheme.

Fig. 6 shows the coded BER performance over the Vehicular B channel with the mobile speed of 30 km/h, which emulates the doubly-selective fading channel. Unlike the conventional large-scale MIMO scheme [8] whose performance degrades about 1.0 dB at the BER of 10^{-4} when mobile speed is increased from 5 km/h to 30 km/h, only 0.1 dB penalty will be paid by the proposed scheme under the same condition. This is caused by that [8] requires the quasi-static channel over the entire transmission frame of large duration, while our proposal could support the application scenarios where the channel is varying not only within the large-size frame, but also within in the small-size subframe. Therefore, we can conclude that the proposed TFT-OFDM scheme for large-scale

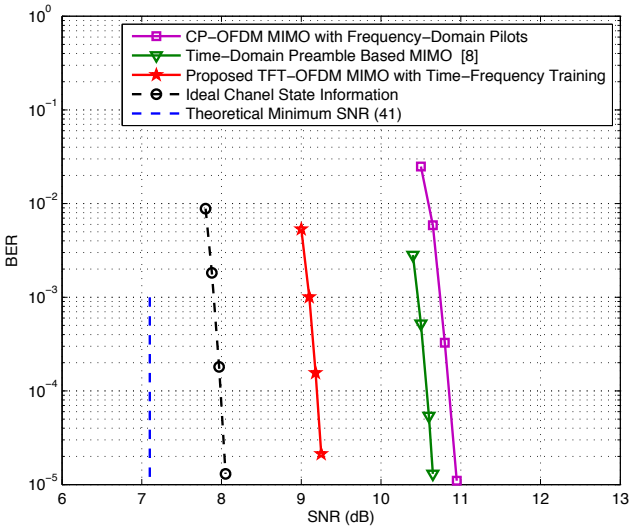

Fig. 5. BER performance comparison between the proposed TFT-OFDM MIMO scheme and its counterparts over the Brazil D channel with the receiver velocity of 5 km/h.

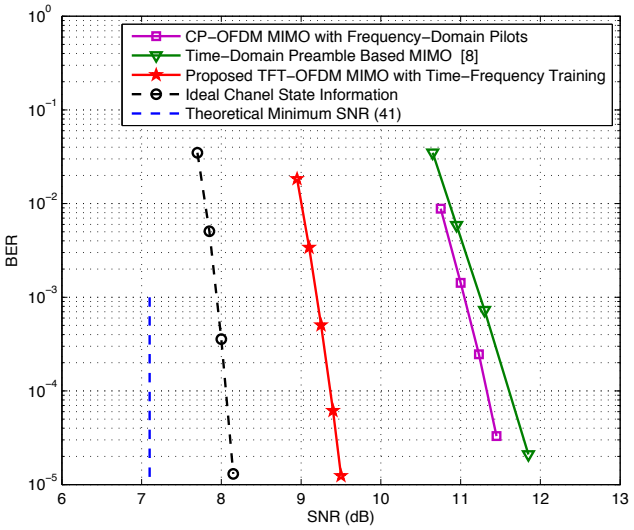

Fig. 6. BER performance comparison between the proposed TFT-OFDM MIMO scheme and its counterparts over the Vehicular B channel with the receiver velocity of 30 km/h.

MIMO systems is robust to fast channel variation, and more obvious performance gain could be expected when the channel is varying faster.

The proposed TFT-OFDM MIMO scheme has superior performance to CP-OFDM MIMO because the time-frequency joint channel estimation could achieve the complete CIR information with high accuracy, while the pilot-based CFR estimation in CP-OFDM MIMO would incur some errors due to interpolation, especially when the channel is deeply frequency-selective and only a small amount of frequency-domain pilots can be used in large-scale MIMO systems. The reason for the performance gain over the iterative channel estimation/data detection scheme in [8] because the latter one assumes quasi-static channel during the entire transmission frame, which deviates a lot from the actual mobile channels. In contrast to the result that CP-OFDM with comb-type pilots performs worse than the time-domain preamble based iterative channel estimation/data detection scheme when the mobile speed is low (e.g., 5 km/h in Fig. 5), CP-OFDM outperforms the preamble based scheme [8] because the frequency-domain pilots within every OFDM symbol could update the CSI more frequently than the time-domain preamble based scheme when the channel is varying fast (e.g., 30 km/h in Fig. 6). Since the spectral efficiency of TFT-OFDM MIMO is higher than its counterparts as addressed in Section V-A, the proposed scheme achieves high spectral efficiency as well as reliable performance due to the time-frequency joint processing.

VII. Conclusions

In this paper, we propose the spectrally efficient TFT-OFDM transmission scheme for large-scale MIMO systems to solve the high-dimensional channel estimation issue in mobile environments. Without cyclic prefix, TFT-OFDM has training information in both the time and the frequency domains for every OFDM symbol, and the frequency-domain grouped pilots occupy much fewer subcarriers than that in common CP-OFDM MIMO systems. This is achieved by the time-frequency joint channel estimation method, whereby the path delays are firstly acquired by the time-domain received TSs without interference cancellation, then there remains much fewer channel parameters to be estimated by fewer frequency-domain pilots. The transmission frame structure composed of one preamble and the subsequent TFT-OFDM symbols could provide efficient channel tracking and data detection in MIMO systems. This paper proves the unification of the time- and frequency-domain channel estimation based on the preamble, and derives CRLB of proposed time-frequency joint channel estimation. Simulation results indicate that the proposed scheme enjoys the BER performance close to the theoretical ergodic capacity. The proposed TFT-OFDM MIMO scheme can be also directly applied in multiple access systems in both the uplink and downlink, and the principle of joint time-frequency processing behind TFT-OFDM can be adapted for other OFDM MIMO systems (including large- and small-scale systems) to achieve higher spectral efficiency as well as more reliable performance over severe fading channels.

Acknowledgments

The first author would like to thank Prof. Feifei Gao for his valuable support for this paper. The authors would like to thank the Guest Editor and the anonymous reviewers for their helpful comments and suggestions to improve the quality of this manuscript.

Appendix Proof of (15) and (16)

We rewrite (14) as below:

$$\mathrm{MSE} = \sigma^2 \mathrm{tr}\left\{\left(\mathbf{c}_0{}^H \mathbf{c}_0\right)^{-1}\right\}. \tag{42}$$

Let $\mathbf{A} = \left(\mathbf{c}_0{}^H \mathbf{c}_0\right)^{-1}$, since $\mathbf{c}_0 = \left[\mathbf{c}_0^{(1)}, \mathbf{c}_0^{(2)}, \cdots, \mathbf{c}_0^{(N_t)}\right]$, the $N_p \times N_p$ matrix $\mathbf{A}$ could be denoted as[7]

$$\mathbf{A} = \begin{bmatrix} \left(\mathbf{c}_0^{(1)}\right)^H \mathbf{c}_0^{(1)} & \left(\mathbf{c}_0^{(1)}\right)^H \mathbf{c}_0^{(2)} & \cdots & \left(\mathbf{c}_0^{(1)}\right)^H \mathbf{c}_0^{(N_t)} \\ \left(\mathbf{c}_0^{(2)}\right)^H \mathbf{c}_0^{(1)} & \left(\mathbf{c}_0^{(2)}\right)^H \mathbf{c}_0^{(2)} & \cdots & \left(\mathbf{c}_0^{(2)}\right)^H \mathbf{c}_0^{(N_t)} \\ \vdots & \vdots & \ddots & \vdots \\ \left(\mathbf{c}_0^{(N_t)}\right)^H \mathbf{c}_0^{(1)} & \left(\mathbf{c}_0^{(N_t)}\right)^H \mathbf{c}_0^{(2)} & \cdots & \left(\mathbf{c}_0^{(N_t)}\right)^H \mathbf{c}_0^{(N_t)} \end{bmatrix}^{-1}_{N_p \times N_p}$$

By denoting the nth diagonal entry of $\mathbf{A}$ as $A(n)$, (42) can be further expressed as

$$\mathrm{MSE} = \sigma^2 \sum_{n=1}^{N_p} A(n) \geq \sigma^2 N_p \sqrt[N_p]{\prod_{n=1}^{N_p} A(n)}, \qquad (43)$$

where the equality holds if and only if all the diagonal entries of $\mathbf{A}$ are identical, i.e.,

$$A(1) = A(2) = \cdots = A(N_p). \qquad (44)$$

According to the Hadamard inequality [50], we have

$$\prod_{n=1}^{N_p} A(n) \geq \det\{\mathbf{A}\}, \qquad (45)$$

where the equality holds if and only if $\mathbf{A}$ is a diagonal matrix, which indicates that $\mathbf{A}^{-1}$ is also a diagonal matrix, i.e., the off-diagonal elements of $\mathbf{A}$ should be zero:

$$\left(\mathbf{c}_0^{(i)}\right)^H \mathbf{c}_0^{(j)} = \mathbf{0}_M, \quad i \neq j. \qquad (46)$$

Due to the perfect autocorrelation property of the Zadoff-Chu sequence [19], we have

$$\left(\mathbf{c}_0^{(i)}\right)^H \mathbf{c}_0^{(i)} = N_p \mathbf{I}_L, \quad 1 \leq i \leq N_t. \qquad (47)$$

Combining (46) and (47), we have

$$\mathbf{c}_0{}^H \mathbf{c}_0 = \mathbf{A}^{-1} = N_p \mathbf{I}_{N_p} = N_t L \mathbf{I}_{N_t L}. \qquad (48)$$

So the optimal design criterion (15) is proved.

Again, by using the Hadamard inequality, we have

$$\det\{\mathbf{A}^{-1}\} = (\det\{\mathbf{A}\})^{-1} \leq {N_p}^{N_p}. \qquad (49)$$

Thus,

$$\det\{\mathbf{A}\} \geq \left(\frac{1}{N_p}\right)^{N_p}, \qquad (50)$$

where the equality holds if and only if $\mathbf{A}$ is a diagonal matrix.

Substituting (45) and (50) into (43), we finally obtain the minimum MSE of the channel estimator as below

$$\mathrm{MSE}_{\min} = \sigma^2 \mathrm{tr}\left\{\frac{1}{N_t L} \mathbf{I}_{N_t L}\right\} = \sigma^2. \qquad (51)$$

So (16) is proved. ■

One remark to make is that, the optimal preamble design criterion (15) and the corresponding minimum MSE (16) are consistent with the results in [13], but the proof technique in this paper differs from that in [13]. Here we acquire the optimal design criterion based on minimizing the MSE, and the Hadamard inequality is used to avoid the inversion of the matrix with large size, while [13] obtained the optimal design criterion based on maximizing the mutual information, and the arithmetic-harmonic means inequality are exploited for the proof.

[7]Note that $N_p = N_t M = N_t L$ as described in Section II.

REFERENCES

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

[2] L. Dai, Z. Wang, and Z. Yang, "Next-generation digital television terrestrial broadcasting systems: Key technologies and research trends," *IEEE Commun. Mag.*, vol. 50, no. 6, pp. 150–158, Jun. 2012.

[3] *IEEE Standard for Information Technology-Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment: Enhancements for Higher Throughput*. IEEE Standard 802.11n-2009, Oct. 2009.

[4] *IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Broadband Wireless Access Systems*. IEEE Standard 802.16-2009, May 2009.

[5] *3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2 (Release 8)*. 3GPP TS 36.300, V8.5.0, May 2008.

[6] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, 2012. (in press).

[7] M. Matthaiou, N. Chatzidiamantis, and G. Karagiannidis, "A new lower bound on the ergodic capacity of distributed MIMO systems," *IEEE Signal Process. Lett.*, vol. 18, no. 4, pp. 227–230, Apr. 2011.

[8] S. K. Mohammed, A. Zaki, A. Chockalingam, and B. S. Rajan, "High-rate space–time coded large-MIMO systems: Low-complexity detection and channel estimation," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 6, pp. 958–974, Dec. 2009.

[9] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[10] H. Taoka and K. Higuchi, "Experiments on peak spectral efficiency of 50 bps/Hz with 12-by-12 MIMO multiplexing for future broadband packet radio access," in *4th IEEE International Symposium on Communications, Control and Signal Processing (ISCCSP'10)*, 2010, pp. 1–6.

[11] G. Breit and et al., *802.11 ac Channel Modeling*. IEEE 802.11 TGac: Very High Throughput below 6GHz Group doc. 09/0088, Jan. 2009.

[12] J. Koivunen, "Characterisation of MIMO propagation channel in multi-link scenarios," *MS Thesis, Helsinki University of Technology, Helsinki, Finland*, Dec. 2007.

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

[14] J. Ahn, H. Lee, and K. Kim, "A near-ML decoding with improved complexity over wider ranges of SNR and system dimension in MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 1, pp. 33–37, Jan. 2012.

[15] C. Knievel, M. Noemm, and P. Hoeher, "Low-complexity receiver for large-MIMO space-time coded systems," in *IEEE Vehicular Technology Conference (VTC Fall)*, Sep. 2011, pp. 1–5.

[16] T. Datta, N. Srinidhi, A. Chockalingam, and B. S. Rajan, "Random-restart reactive tabu search algorithm for detection in large-MIMO systems," *IEEE Commun. Lett.*, vol. 14, no. 12, pp. 1107–1109, Dec. 2010.

[17] P. Som, T. Datta, N. Srinidhi, A. Chockalingam, and B. S. Rajan, "Low-complexity detection in large-dimension MIMO-ISI channels using graphical models," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 8, pp. 1497–1511, Dec. 2011.

[18] O. Rousseaux, G. Leus, P. Stoica, and M. Moonen, "Gaussian maximum-likelihood channel estimation with short training sequences," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2945–2955, Nov. 2005.

[19] L. Bomer and M. Antweiler, "Perfect N-phase sequences and arrays," *IEEE J. Sel. Areas Commun.*, vol. 10, no. 4, pp. 782–789, May 1992.

[20] D. Sarwate, "Bounds on crosscorrelation and autocorrelation of sequences," *IEEE Trans. Inf. Theory*, vol. 25, no. 6, pp. 720–724, Nov. 1979.

[21] L. Dai, Z. Wang, and S. Chen, "A novel uplink multiple access scheme based on TDS-FDMA," *IEEE Trans. Wireless Commun.*, vol. 10, no. 3, pp. 757–761, Mar. 2011.

[22] L. Dai, Z. Wang, J. Wang, and Z. Yang, "Joint channel estimation and time-frequency synchronization for uplink TDS-OFDMA systems," *IEEE Trans. Consum. Electron.*, vol. 56, no. 2, pp. 494–500, May 2010.
[23] H. Minn and N. Al-Dhahir, "Optimal training signals for MIMO OFDM channel estimation," *IEEE Trans. Wireless Commun.*, vol. 5, no. 5, pp. 1158–1168, May 2006.
[24] Z. Tang, R. Cannizzaro, G. Leus, and P. Banelli, "Pilot-assisted time-varying channel estimation for OFDM systems," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 2226–2238, May 2007.
[25] J. Haupt, W. Bajwa, G. Raz, and R. Nowak, "Toeplitz compressed sensing matrices with applications to sparse channel estimation," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5862–5875, Nov. 2010.
[26] M. Duarte and Y. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4053–4085, Sep. 2011.
[27] *Guideline for Evaluation of Radio Transmission Technology for IMT-2000*. Recommendation ITU-R M. 1225, 1997.
[28] I. Telatar and D. Tse, "Capacity and mutual information of wideband multipath fading channels," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1384–1400, Jul. 2000.
[29] T. Sarkar, Z. Ji, K. Kim, A. Medouri, and M. Salazar-Palma, "A survey of various propagation models for mobile communication," *IEEE Antennas Propagat. Mag.*, vol. 45, no. 3, pp. 51–82, Mar. 2003.
[30] *Digital Video Broadcasting (DVB); Frame Structure, Channel Coding and Modulation for a Second Generation Digital Terrestrial Television Broadcasting System (DVB-T2)*. ETSI Standard, EN 302 755, V1.3.1, Apr. 2012.
[31] *Digital Television Systems-Brazilian Tests-Final Report*. SET/ABERT ANATEL SP, May 2000.
[32] X. Wang, H. Li, and H. Lin, "A new adaptive OFDM system with precoded cyclic prefix for dynamic cognitive radio communications," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 2, pp. 431–442, Feb. 2011.
[33] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volumn I: Estimation Theory*. New Jersey, USA: Prentice-Hall, 1993.
[34] B. Muquet, Z. Wang, G. Giannakis, M. De Courville, and P. Duhamel, "Cyclic prefixing or zero padding for wireless multicarrier transmissions?" *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 2136–2148, Dec. 2002.
[35] S. Ohno and G. Giannakis, "Optimal training and redundant precoding for block transmissions with application to wireless OFDM," *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 2113–2123, Dec. 2002.
[36] J. Wang, Z. Yang, C. Pan, and J. Song, "Iterative padding subtraction of the PN sequence for the TDS-OFDM over broadcast channels," *IEEE Trans. Consum. Electron.*, vol. 51, no. 11, pp. 1148–1152, Nov. 2005.
[37] C. Komninakis, C. Fragouli, A. H. Sayed, and R. D. Wesel, "Multi-input multi-output fading channel tracking and equalization using Kalman estimation," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1065–1076, May 2002.
[38] F. Wan, W. Zhu, and M. Swamy, "Semi-blind most significant tap detection for sparse channel estimation of OFDM systems," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 57, no. 3, pp. 703–713, Mar. 2010.
[39] K. Vishnu Vardhan, S. K. Mohammed, A. Chockalingam, and B. S. Rajan, "A low-complexity detector for large MIMO systems and multicarrier CDMA systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 3, pp. 473–485, Apr. 2008.
[40] P. Som, T. Datta, A. Chockalingam, and B. Rajan, "Improved large-MIMO detection based on damped belief propagation," in *IEEE Information Theory Workshop (ITW'10)*, Jan. 2010, pp. 1–5.
[41] M. Suneel, P. Som, A. Chockalingam, and B. S. Rajan, "Belief propagation based decoding of large non-orthogonal STBCs," in *IEEE International Symposium on Information Theory (ISIT'09)*, Jul. 2009, pp. 2003–2007.
[42] S. Yang, T. Lv, R. Maunder, and L. Hanzo, "Unified bit-based probabilistic data association aided MIMO detection for high-order QAM constellations," *IEEE Trans. Veh. Technol.*, vol. 60, no. 3, pp. 981–991, Mar. 2011.
[43] Z. Yang, L. Dai, J. Wang, J. Wang, and Z. Wang, "Transmit diversity for TDS-OFDM broadcasting system over doubly selective fading channels," *IEEE Trans. Broadcast.*, vol. 57, no. 1, pp. 135–142, Mar. 2011.
[44] S. Tang, F. Yang, K. Peng, C. Pan, K. Gong, and Z. Yang, "Iterative channel estimation for block transmission with known symbol padding - a new look at TDS-OFDM," in *IEEE Global Telecommunications Conference (GLOBECOM'07)*, Washington, DC, May 2007, pp. 4269–4273.
[45] S. Sesia, I. Toufik, and M. Baker, *LTE-The UMTS Long Term Evolution: From Theory to Practice*. New Jersey, SA: John Wiley & Sons, 2009.
[46] X. Liang, "An algebraic, analytic, and algorithmic investigation on the capacity and capacity-achieving input probability distributions of finite-input–finite-output discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 1003–1023, Jun. 2008.
[47] P. Elia, B. Sethuraman, and P. Vijay Kumar, "Perfect space–time codes for any number of antennas," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 3853–3868, Nov. 2007.
[48] T. Richardson, M. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 619–637, Feb. 2001.
[49] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
[50] C. Li, Y. Lin, S. Tsai, and P. Vaidyanathan, "Optimization of transceivers with bit allocation to maximize bit rate for MIMO transmission," *IEEE Trans. Commun.*, vol. 57, no. 12, pp. 3556–3560, Dec. 2009.

Linglong Dai (M'11) received his B.S. degree from Zhejiang University in 2003, M.S. degree (with the highest honor) from China Academy of Telecommunications Technology (CATT) in 2006, and Ph.D. degree (with the highest honor) from Tsinghua University in 2011, respectively. He is now a Post Doctoral Fellow with the Department of Electronic Engineering, Tsinghua University, Beijing, China. His research focuses on wireless and optical communications. He has published over 20 journal and conference papers. He was awarded the 2011 Tsinghua Excellent Doctor of Electronic Engineering, the 2011 Tsinghua Academic Star, and the 2012 Beijing Excellent Doctoral Dissertations.

Zhaocheng Wang (M'06-SM'11) received his B.S., M.S. and Ph.D. degrees from Tsinghua University in 1991, 1993 and 1996, respectively. From 1996 to 1997, he was with Nanyang Technological University (NTU) in Singapore as a Post Doctoral Fellow. From 1997 to 1999, he was with OKI Techno Centre (Singapore) Pte. Ltd., firstly as a research engineer and then as a senior engineer. From 1999 to 2009, he worked at SONY Deutschland GmbH, firstly as a senior engineer and then as a principal engineer. He is currently a Professor at the Department of Electronic Engineering, Tsinghua University. His research areas include wireless communications, digital broadcasting and millimeter wave communications. He holds 25 granted US/EU patents and has published over 70 technical papers. He has served as technical program committee co-chair/member of many international conferences. He is a Senior Member of IEEE and a Fellow of IET.

Zhixing Yang (M'00-SM'11) received his B.S. degree from the Department of Electronic Engineering, Tsinghua University, Beijing, China, in 1970. He is now a full professor at the Department of Electronics Engineering of Tsinghua University, Beijing, China. He is the executive director of the State Key Laboratory on Microwave and Digital Communications, China, and the executive director of the development group of the digital television terrestrial broadcasting state standard for China. He received several national awards and held dozens of patents. His research interests are in high-speed data transmission over broadband digital television terrestrial broadcasting, wireless links, wireless communication theory and communication systems design. He is a Senior Member of IEEE.