Design and Optimization on Training Sequence for mmWave Communications: A New Approach for Sparse Channel Estimation in Massive MIMO

Xu Ma, *Student Member, IEEE*, Fang Yang, *Senior Member, IEEE*, Sicong Liu, *Student Member, IEEE*, Jian Song, *Fellow, IEEE*, and Zhu Han, *Fellow, IEEE*

***Abstract*— In the next generation of cellular networks, millimeter wave (mmWave) communications will play an important role. With the utilization of mmWave communications, the massive multiple-input multiple-output (MIMO) technique can be effectively employed, which will significantly improve system capacity. However, an effective channel estimation scheme is the prerequisite of system stability and in great need for improvement in massive MIMO systems. In this paper, a channel estimation scheme based on training sequence (TS) design and optimization with high accuracy and spectral efficiency is investigated in the framework of structured compressive sensing. As a new perspective to optimize the block coherence of the sensing matrix, the auto-coherence and cross-coherence of the blocks are proposed and specified as two kinds of key merit factors. In order to optimize the two factors, specific TS is designed and obtained from the inverse discrete Fourier transform of a frequency domain binary training sequence, and a genetic algorithm is adopted afterwards to optimize the merit factors of the TS. It is demonstrated by the simulation results that the block coherence of the sensing matrix can be significantly reduced by the proposed TS design and optimization method. Moreover, by using the proposed optimized TS's, the channel estimation outperforms the conventional TS design obtained by the brute force search in terms of the correct recovery probability, mean square error, and bit error rate, and can also approach the Cramer–Rao lower bound.**

Manuscript received October 27, 2016; revised February 21, 2017; accepted March 15, 2017. Date of publication April 27, 2017; date of current version June 19, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 61401248, in part by the New Generation Broadband Wireless Mobile Communication Network of the National Science and Technology Major Projects under Grant 2015ZX03002008, in part by the Tsinghua University Initiative Scientific Research Program under Grant 2014Z06098, in part by the International Cooperation of Collaborative Innovation in Science and Technology Project of Shenzhen, in part by the Young Elite Scientist Sponsorship Program by CAST, and in part by the U.S. NSF under Grant CPS-1646607, Grant ECCS-1547201, Grant CCF-1456921, Grant CNS-1443917, and Grant ECCS-1405121. *(Corresponding author: Fang Yang.)*

X. Ma and S. Liu are with the Electronic Engineering Department, Tsinghua University, Beijing 100084, China, and also with the Tsinghua National Laboratory for Information Science and Technology, Research Institute of Information Technology, Tsinghua University, Beijing 100084, China (e-mail: maxu14@mails.tsinghua.edu.cn; liu-sc12@mails.tsinghua.edu.cn).

F. Yang and J. Song are with the Electronic Engineering Department, Tsinghua University, Beijing 100084, China, also with the Tsinghua National Laboratory for Information Science and Technology, Research Institute of Information Technology, Tsinghua University, Beijing 100084, China, and also with the Shenzhen City Key Laboratory of Digital TV System, Shenzhen 518057, China (e-mail: fangyang@tsinghua.edu.cn; jsong@tsinghua.edu.cn).

Z. Han is with the Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77204 USA (e-mail: hanzhu22@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSAC.2017.2698978

***Index Terms*— Millimeter wave communications, massive MIMO, channel estimation, training sequence, structured compressive sensing (SCS).**

I. Introduction

IN THE future 5G communications, a significantly large number of applications will be supported compared with 4G communications [1]. Therefore, the demand of data capacity will dramatically increased. One important resolution of this problem is to improve the bandwidth. As one of the most effective and innovative solutions to realize the aforementioned 5G vision and requirements, the use of a large band of spectrum in the very high frequencies such as the millimeter wave (mmWave) bands has recently gained significant interests. The utilization of mmWave communications [2]–[7], which are applied under the frequency of 30-300GHz corresponding to wavelengths from 1 to 10 mm, can provide an extremely broad bandwidth that none of the existence cellular communications have ever seen. MmWave also has the advantage of limited inter-cell interference, low transmission latency, and improved security. Therefore, it has already been applied in commercial products by standards like IEEE 802.11ad [8], and thus as an potential technique in the future 5G communications [5], [9].

The history of the mmWave technology began in 1895, when an Indian physicist J. C. Bose demonstrated wireless signaling at frequencies as high as 60GHz [10]. In 1947, J. H. Van Vleck observed that the oxygen molecule absorbs electromagnetic energy more significantly at 60-GHz than at other frequencies [11], whose investigation and study would be mainly driven by military and space applications during the 1960s to 1980s [12]. Later when microcellular communication became the major driver from the late 1980s to the early 1990s, people tried to limit the link distance and allow for higher frequency reuse to increase the network capacity [13]. In the mid-1990s, it was realized that unlicensed use could be an appropriate regime for using mmWave spectrum. The first 60-GHz regulations was issued by Japan for unlicensed utilization in the $59-66$ GHz band in 2000, whose maximum transmit power is limited to 10 dBm with the maximum transmission bandwidth of 2.5 GHz [14]. After then, a number of countries such as the United States, Canada, and Europe issued their mmWave regulations. Now, the mmWave technique has become an important candidate for future 5G communications.

With the utilization of mmWave communications, massive multiple input multiple output (MIMO) is available to be applied in the future 5G communications [15]. Massive MIMO has become one of the key techniques in communication systems with the rapid development of the wireless cellular network technology [16]–[19], mainly due to its high spectral and energy efficiency. Meanwhile, orthogonal frequency division multiplexing (OFDM) can also improve system performance and increase spectral efficiency under frequency-selective channels. Therefore, integration of both massive MIMO and OFDM technologies, which is usually denoted as massive MIMO-OFDM, is essential to enhance system capacity. Massive MIMO-OFDM is becoming a promising alternative among the upcoming 5G technologies.

As the modern cellular networks aim at providing a high data rate, accurate channel estimation is critical [20], especially for the massive MIMO-OFDM system with the utilization of mmWave communications. To accurately recover the data, each receiver has to acquire multiple channel state information (CSI) whose amount is equal to the amount of the transmit antennas. As a result, the complexity and resource consumption in a massive MIMO-OFDM system are usually much more complicated than those in a single input single output (SISO) system. The state-of-the-art channel estimation schemes are usually implemented by exploiting time-domain training sequences (TS's) [21] or frequency-domain pilots [22]. However, in the conventional orthogonal channel estimation schemes, the overhead will also increase with the number of transmit antennas, inevitably causing significant degradation in spectral efficiency.

Fortunately, it is possible to recover a sparse signal through only a small amount of observations by utilizing the compressive sensing (CS) theory [23]–[28], which is an effective method to improve spectral efficiency. Since the number of distinguishable multipath delays are much smaller compared with the channel delay spread, the channel impulse response (CIR) of the wireless channel is usually a sparse vector [20]. Therefore, CS theory can be applied to channel estimation naturally. Some work has been done in the CS-based channel estimation [29]–[35]. Time-domain TS was adopted to perform SISO-OFDM channel estimation [29], [30]. A non-orthogonal frequency-domain pilot scheme was proposed [31], while an approach using frequency-domain pilot with the time-domain TS as an auxiliary information was investigated [32]. For mmWave channel estimation, literature [33] proposed a novel hierarchical multi-resolution codebook to perform channel estimation, while a distributed compressive sensing based channel estimation scheme was proposed in [34]. A fast channel estimation (FCE) and rate-adaptive channel estimation (RACE) algorithm were proposed in [35]. These works assumed that the mmWave channels exhibit the angle-domain sparsity because the path loss for non-line-of-sight (NLOS) is much higher than that for line-of-sight (LOS). However time-domain training sequence based channel estimation scheme is not considered in those literatures. The TS optimization is worthy to be addressed, which is very important to improve system performance, and thus requires thorough investigation.

In this paper, a novel channel estimation scheme with time-domain TS design and optimization is proposed for a MIMO-OFDM based mmWave system to solve these problems. Numerical simulation results demonstrate that the proposed TS design and optimization method can significantly reduce the block coherence of the sensing matrix. Furthermore, the optimized TS has a better channel estimation performance than the counterparts, which is reflected in the correct recovery probability, mean square error (MSE), and bit error rate (BER), and can approach the Cramer-Rao lower bound (CRLB). The contributions of this paper is summarized as follows.

- A novel approach of channel estimation based on TS is proposed. The receiver takes advantage of the inter block interference free (IBI-free) region [30] in the TS between OFDM data blocks to perform the sparse channel estimation. Using the knowledge of the spatial correlation among different channels from the transmit antennas, a block sparse model can be adopted and the structured compressive sensing (SCS) framework is built, which effectively increases spectral efficiency.
- The auto-coherence and cross-coherence of the blocks are proposed as two key merit factors for block coherence optimization, which brings a new perspective to optimize the block coherence of the sensing matrix. It is explained that minimizing the auto-coherence and cross-coherence, which is more simple and convenient, is equivalent to minimizing the block coherence.
- To optimize the auto-coherence, the inverse discrete Fourier transform (IDFT) TS pattern with the cyclic extension structure is adopted, which has a relatively small value of auto-coherence in nature. Moreover, the genetic algorithm (GA) [36] is applied to perform the cross-coherence optimization, which can acquire a small value of cross-coherence.

The remainder of this paper is organized as follows. The wireless channel model and the massive MIMO-OFDM system model are introduced in Section II. The proposed TS design and optimization criterion based on SCS is investigated in Section III. The proposed auto-coherence and cross-coherence optimization are investigated in Sections IV and V, respectively. Simulation results and discussion are addressed in Section VI. Finally, we draw conclusions in Section VII.

Notation: Uppercase and lowercase boldface letters are adopted to denote matrices and column vectors; $(\cdot)^T$, $(\cdot)^H$, $\|\cdot\|_p$, $\mathrm{E}(\cdot)$, and $\boldsymbol{A}(i)$ denote transpose, conjugate transpose, l_p norm operations, the expectation of a random variable, and the i-th column of the matrix or the i-th element of a vector, respectively.

II. System Model

In this paper, the TS design and optimization scheme are mainly applied in the time-domain synchronous OFDM (TDS-OFDM) systems [37], whose guard interval can be adopted as the TS. The OFDM frame $\boldsymbol{t}$ is composed of the guard interval $\boldsymbol{c}$ and the OFDM data block $\boldsymbol{x}$, which can be represented as $\boldsymbol{t} = [\boldsymbol{c}^T, \boldsymbol{x}^T]^T$. Apart from protection of the data block, the guard interval can also be padded by the TS, which is

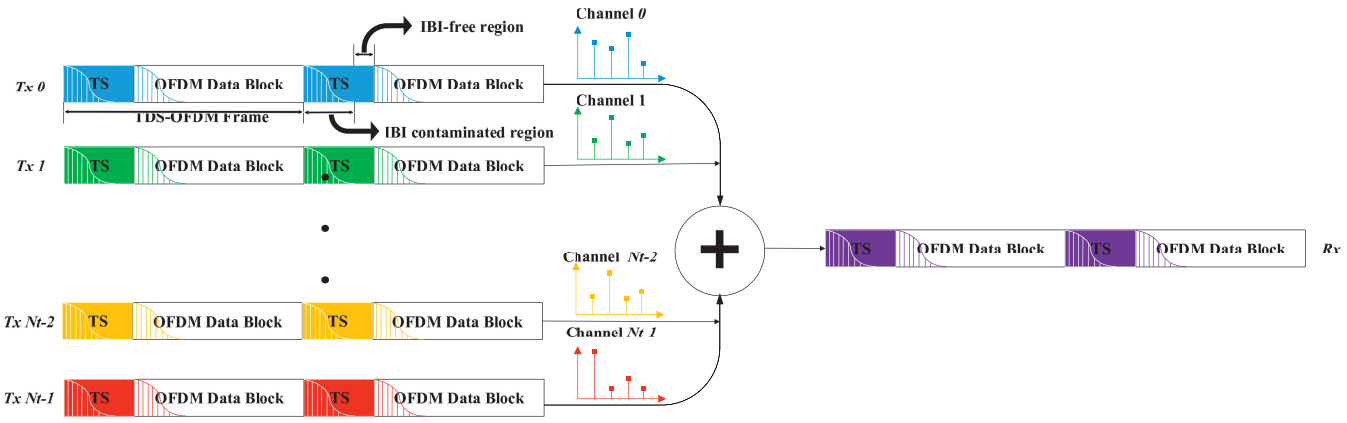

Fig. 1. The structure of the proposed massive MIMO-OFDM frames.

used for channel estimation in this paper. For a MIMO system with N_t transmit antennas and N_r receive antennas, in order to distinguish different OFDM frames from different transmit antennas, an additional subscript is used to describe the OFDM frame from the i-th transmit antenna, i.e.,

$$\boldsymbol{t}_i = [\boldsymbol{c}_i^T, \boldsymbol{x}_i^T]^T, \tag{1}$$

where $0 \leq i < N_t$, $\boldsymbol{c}_i = \left[c_{i,0}, c_{i,1}, \ldots, c_{i,M-1}\right]^T$ and $\boldsymbol{x}_i = \left[x_{i,0}, x_{i,1}, \ldots, x_{i,K-1}\right]^T$. Here $c_{i,j}$ and $x_{i,j}$ are the elements of the M-length TS and K-length data block, respectively.

For mmWave communications, the channel in the delay-domain can be modeled as [38],

$$\boldsymbol{H}(\tau) = \sum_{l=0}^{s-1} \boldsymbol{H}_l \delta(\tau - \tau_l), \tag{2}$$

where τ_l is the delay of the l-th path, s is the number of the multipath, $\boldsymbol{H}, \boldsymbol{H}_l \in C^{N_r \times N_t}$ are the CIR matrix associate with the l-th path, whose elements are CIR from the specific transmit and receive antennas. $\boldsymbol{H}_l$ can be given by

$$\boldsymbol{H}_l = \alpha_l a_R(\theta_l) a_T^*(\varphi_l), \tag{3}$$

where α_l is the complex gain of the l-th path, $\theta_l, \varphi_l \in [0, 2\pi]$ are azimuth angles of arrival (AoA) and angles of departure (AoD). $a_R(\theta_l)$ and $a_T(\varphi_l)$ are steering vectors at the receiver and transmitter, respectively, and can be denoted as

$$\begin{aligned} a_R(\theta_l) &= \left[1, e^{j2\pi d \sin(\theta_l)/\lambda}, \ldots, e^{j2\pi(N_t-1)d\sin(\theta_l)/\lambda}\right]^T, \\ a_T(\varphi_l) &= \left[1, e^{j2\pi d \sin(\varphi_l)/\lambda}, \ldots, e^{j2\pi(N_r-1)d\sin(\varphi_l)/\lambda}\right]^T, \end{aligned} \tag{4}$$

where λ and d are wavelength and antenna spacing, respectively. These formulas are valid when the uniform linear arrays (ULAs) are assumed for the antennas, which is utilized in the simulations.

Equation (2) is the CIR expression and relation among different transmit and receive antennas. However, this model may face a severe problem of power leakage caused by the continuous angels of arrival or departure [34]. Therefore, a more general channel model is utilized in this paper. For the channel from one transmit and receive antenna, the CIR

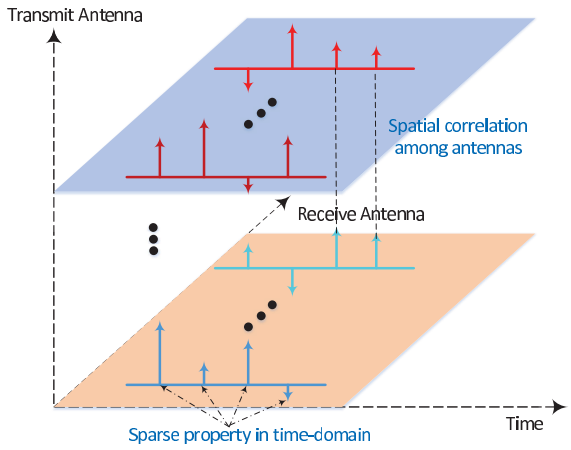

Fig. 2. The structure of the proposed massive MIMO-OFDM frames.

associated with the i-th ($0 \leq i < N_t$) transmit antenna can be modeled as

$$\boldsymbol{h}_i = \left[h_{i,0}, h_{i,1}, \ldots, h_{i,L-1}\right]^T, \tag{5}$$

where L is the channel length. As the processing at each receiver after passing through the channel is identical for every receive antenna, the receive antenna index is omitted in this paper. It is widely known that the CIR of the wireless channel can be modeled as a sparse vector [39]. Besides, according to (2), only s multipaths are non-zero in the delay domain in the mmWave channel. In other words, the number of the non-zero elements s in vector $\boldsymbol{h}_i$ is much less than the channel length L, e.g., $s \ll L$, which is shown in Fig. 2. The non-zero support of $\boldsymbol{h}_i$ can be represented as $T_i = \left\{l : \left|h_{i,l}\right| > 0\right\}_{l=0}^{L-1}$.

Through the channel, the received signal $\mathbf{r}_i$ from the i-th transmit antenna is the convolution of the transmit signal and the CIR vector, i.e.,

$$\boldsymbol{r}_i = \boldsymbol{t}_i * \boldsymbol{h}_i + \boldsymbol{n}_i, \tag{6}$$

where $\mathbf{n}_i$ denotes the additive white Gaussian noise (AWGN) vector in which each element is subject to a Gaussian independent and identical distribution (i.i.d.) with zero mean and variance of σ^2. For channel estimation, we only focus on

the TS part within the received signal $\mathbf{r}_{\mathrm{TS},i}$ and (6) can be rewritten as

$$\boldsymbol{r}_{\mathrm{TS},i} = \boldsymbol{\Psi}_i \boldsymbol{h}_i + \boldsymbol{n}_{\mathrm{TS},i}, \tag{7}$$

where $\boldsymbol{n}_{\mathrm{TS},i}$ is the corresponding noise vector. $\boldsymbol{\Psi}_i$ is a Toeplitz matrix determined by the previous OFDM data block and the current TS, which can be denoted as

$$\boldsymbol{\Psi}_i = \begin{bmatrix} c_{i,0} & x'_{i,K-1} & x'_{i,K-2} & \cdots & x'_{i,K-L+1} \\ c_{i,1} & c_{i,0} & x'_{i,K-1} & \cdots & x'_{i,K-L+2} \\ c_{i,2} & c_{i,1} & c_{i,0} & \cdots & x'_{i,K-L+3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{i,L-1} & c_{i,L-2} & c_{i,L-3} & \cdots & c_{i,0} \\ c_{i,L} & c_{i,L-1} & c_{i,L-2} & \cdots & c_{i,1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{i,M-1} & c_{i,M-2} & c_{i,M-3} & \cdots & c_{i,M-L} \end{bmatrix}_{M\times L}, \tag{8}$$

where $x'_{i,k}$ is the k-th element of the previous OFDM data block.

When the TS is used to perform the channel estimation, there is a significant challenge that the TS is contaminated by the previous OFDM data block, whose influence is significant under multipath channels with long channel delay. In other words, sensing matrix $\boldsymbol{\Psi}_i$ has the elements of the previous OFDM data block as illustrated in (8), which is not available for CIR recovery. A new approach utilizing the IBI-free region to perform channel estimation under the framework of CS has been proposed in [30]. The IBI-free region is the rear part of the TS, which is immune from the influence of the previous OFDM data block as shown in Fig. 1 intuitively. It should be noted that only one receive antenna is shown in the figure because the receive antennas is independent with each other.

In this way, the received signal in the IBI-free region can be represented as

$$\widetilde{\boldsymbol{r}}_i = \widetilde{\boldsymbol{\Psi}}_i \boldsymbol{h}_i + \widetilde{\boldsymbol{n}}_i, \tag{9}$$

where $\widetilde{\boldsymbol{r}}_i$ and $\boldsymbol{h}_i$ are the received TS in the IBI-free region and the CIR vector from the i-th transmit antenna, respectively. $\widetilde{\boldsymbol{n}}_i$ is the AWGN vector, in which each element is subject to an i.i.d. Gaussian distribution with zero mean, variance σ^2, and size of $M - L$. The sensing matrix $\widetilde{\boldsymbol{\Psi}}_i$ is given by

$$\widetilde{\boldsymbol{\Psi}}_i = \begin{bmatrix} c_{i,L-1} & c_{i,L-2} & c_{i,L-3} & \cdots & c_{i,0} \\ c_{i,L} & c_{i,L-1} & c_{i,L-2} & \cdots & c_{i,1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{i,M-1} & c_{i,M-2} & c_{i,M-3} & \cdots & c_{i,M-L} \end{bmatrix}_{N\times L}, \tag{10}$$

where $N = M - L + 1$. Actually, equation (9) is the sub-equation of (7) with respect to the last N rows.

For each receive antenna, the received signal in the IBI-free region is the superposition of all the signal from different transmit antennas as shown in Fig. 1, which can be represented as,

$$\boldsymbol{r} = \sum_{i=0}^{N_t-1} \boldsymbol{r}_i = \sum_{i=0}^{N_t-1} \boldsymbol{\Psi}_i \boldsymbol{h}_i + \boldsymbol{n} = \boldsymbol{\Psi}\boldsymbol{h} + \boldsymbol{n}, \tag{11}$$

where $\boldsymbol{\Psi}$ and $\boldsymbol{h}$ are made up of sensing matrices and CIR vectors from different transmit antennas, i.e., $\boldsymbol{\Psi} = \left[\boldsymbol{\Psi}_0, \boldsymbol{\Psi}_1, \ldots, \boldsymbol{\Psi}_{N_t-1}\right]$ and $\boldsymbol{h} = \left[\boldsymbol{h}_0^T, \boldsymbol{h}_1^T, \ldots, \boldsymbol{h}_{N_t-1}^T\right]^T$.

The channel model used in this paper is different from that in [33]–[35], whose channels of different antennas will have only phase difference. In the model of this paper, both the amplitude and the phase of the different channel taps can be variant. However, the propagation delay is roughly the same for all transmit-receive antenna pairs, there exists a spatial correlation for the channels among different antennas. The non-zero supports are identical for all the channels within the same receive antenna [40], [41], whereas the non-zero coefficients are different. We rearrange the elements of the CIR vector and extract the taps of the CIR with the identical index from different transmit antennas into a group. Then the received signal in (11) can be rewritten as

$$\boldsymbol{r} = \boldsymbol{\Phi}\boldsymbol{g} + \boldsymbol{n}, \tag{12}$$

where $\boldsymbol{\Phi} = \left[\boldsymbol{\Phi}_0, \boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_{L-1}\right]$ with $\boldsymbol{\Phi}_i = \left[\boldsymbol{\Psi}_0(i), \boldsymbol{\Psi}_1(i), \ldots, \boldsymbol{\Psi}_{N_t-1}(i)\right]$, and $\boldsymbol{g} = \left[\boldsymbol{g}_0^T, \boldsymbol{g}_1^T, \ldots, \boldsymbol{g}_{L-1}^{\mathrm{T}}\right]^{\mathrm{T}}$ with $\boldsymbol{g}_i = \left[\boldsymbol{h}_0(i), \boldsymbol{h}_1(i), \ldots, \boldsymbol{h}_{N_t-1}(i)\right]^T$, where $0 \le i < L$. Considering the identical non-zero support of the different CIR vector, the entries of $\boldsymbol{g}_i$ will be either all zero or all non-zero. In this way, the integrated CIR vector is divided into sparse blocks and the channel estimation can be performed in the framework of SCS which will be investigated in the next section.

III. TS Design and Optimization

In the framework of CS, the CIR recovery performance is related with the sensing matrix $\boldsymbol{\Phi}$, which is determined by the TS's of all transmit antennas. A properly designed TS will improve the performance of the sensing matrix, and thus increase the channel estimation accuracy. Based on CS theory, a significant approach to evaluate the performance of a sensing matrix is the restricted isometry property (RIP) [42]. If there exists a constant $\delta(0 < \delta < 1)$ and for all S-sparse vectors $\boldsymbol{\eta} \in C^L$ satisfying

$$(1-\delta)\left\|\boldsymbol{\eta}\right\|_2^2 \le \left\|\boldsymbol{\Psi}\boldsymbol{\eta}\right\|_2^2 \le (1+\delta)\left\|\boldsymbol{\eta}\right\|_2^2, \tag{13}$$

the sensing matrix $\boldsymbol{\Psi}$ is considered to fulfill RIP. However, there is no polynomial time computational algorithm to identify RIP satisfaction of a matrix [43]. As an alternative evaluation method, the mutual incoherence property (MIP), which is the maximum coherence between different columns, is easy to be calculated and can be a necessary and sufficient condition especially when the greedy algorithms are adopted in the recovery process. The critical merit factor of the sensing matrix, the *coherence* is defined as

$$\mu = \max_{0\le l,k<L, l\ne k} \frac{\left|\langle \boldsymbol{\varphi}_l, \boldsymbol{\varphi}_k \rangle\right|}{\left\|\boldsymbol{\varphi}_l\right\|_2 \cdot \left\|\boldsymbol{\varphi}_k\right\|_2}, \tag{14}$$

where $\boldsymbol{\varphi}_l$ and $\boldsymbol{\varphi}_k$ are the l-th and the k-th column of the sensing matrix, respectively. Many researchers have successfully improved the sparse signal recovery performance by optimizing the coherence of the sensing matrix [44], [45], including our previous work in literature [29].

Extending MIP to SCS, the coherence is inefficient and poorly performed as the column number of the sensing matrix become larger. The SCS framework emphasizes the signal with a block sparse property, which needs to be taken into account for the MIP criterion. Considering the block sparse property, the columns of the sensing matrix with respect to the identical supports should be combined as block sets. The *block coherence*, which is the merit factor in the SCS-MIP framework, should evaluate the coherence among the block sets. Therefore, the so-called block coherence is proposed [46], which is defined as

$$\mu_{\mathrm{B}}(\mathbf{\Phi}) = \max_{0\le l,k<L, l\neq k} \frac{\rho\left(\mathbf{\Phi}_l^H \mathbf{\Phi}_k\right)}{N_t}, \tag{15}$$

where $\rho(\boldsymbol{A})$ is the spectrum norm of a matrix and represents the maximum singular value of matrix $\boldsymbol{A}$. The columns of blocks are normalized for convenience. The block coherence helps analyse the maximum coherence between different blocks of the sensing matrix instead of column coherence, and is adopted to evaluate the sensing matrix in this paper.

The block sparse signal recovery will be simpler when each block of the sensing matrix can be easily distinguished from others. Therefore, a small value of block coherence can decrease the probability of confusing different blocks in the recovery process, which is an important approach to improve the accuracy of recovery performance. Therefore, to obtain a well-performed sensing matrix, optimizing the block coherence is vital, i.e.,

$$\mathbf{\Phi}_{\mathrm{opt}} = \arg\min_{\mathbf{\Phi}} \left(\mu_{\mathrm{B}}(\mathbf{\Phi})\right). \tag{16}$$

For this problem in (16), a closed form solution can hardly be acquired and the brute force search is also impossible as the search space is tremendous. A suboptimal design scheme can be based on two considerations. One is to compress the search space with adding some restricted conditions of the TS's, while the other is to investigate some efficient search algorithms.

We first address a simple case with $N_t = 2$ transmit antennas as an example. Two arbitrary blocks are selected, which are denoted as $\mathbf{\Phi}_\alpha = [\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1]$ and $\mathbf{\Phi}_\beta = [\boldsymbol{\beta}_0, \boldsymbol{\beta}_1]$. The coherence between the two blocks can be represented as

$$\mu_{\mathrm{B}}^2(\mathbf{\Phi}_\alpha, \mathbf{\Phi}_\beta) = \rho^2\left(\mathbf{\Phi}_\alpha^H \mathbf{\Phi}_\beta\right)/4 = \frac{\rho^2\left(\begin{bmatrix} \boldsymbol{\alpha}_0^H\boldsymbol{\beta}_0 & \boldsymbol{\alpha}_0^H\boldsymbol{\beta}_1 \\ \boldsymbol{\alpha}_1^H\boldsymbol{\beta}_0 & \boldsymbol{\alpha}_1^H\boldsymbol{\beta}_1 \end{bmatrix}\right)}{4}, \tag{17}$$

wherein the elements of matrix $\mathbf{\Phi}_\alpha^H \mathbf{\Phi}_\beta$ are the coherences between different columns in $\mathbf{\Phi}_\alpha$ and $\mathbf{\Phi}_\beta$.

The singular values of matrix $\mathbf{\Phi}_\alpha^H \mathbf{\Phi}_\beta$ is equal to the square root of the characteristic values of $(\mathbf{\Phi}_\alpha^H \mathbf{\Phi}_\beta)^H (\mathbf{\Phi}_\alpha^H \mathbf{\Phi}_\beta)$, which can be written as,

$$(\mathbf{\Phi}_\alpha^H \mathbf{\Phi}_\beta)^H (\mathbf{\Phi}_\alpha^H \mathbf{\Phi}_\beta) = \begin{bmatrix} m_{0,0} & m_{0,1} \\ m_{1,0} & m_{1,1} \end{bmatrix}, \tag{18}$$

where

$$\begin{aligned} m_{0,0} &= \left|\boldsymbol{\alpha}_0^H\boldsymbol{\beta}_0\right|^2 + \left|\boldsymbol{\alpha}_1^H\boldsymbol{\beta}_0\right|^2, \\ m_{1,1} &= \left|\boldsymbol{\alpha}_0^H\boldsymbol{\beta}_1\right|^2 + \left|\boldsymbol{\alpha}_1^H\boldsymbol{\beta}_1\right|^2, \\ m_{0,1} &= m_{1,0}^* = \boldsymbol{\beta}_0^H\boldsymbol{\alpha}_0\boldsymbol{\alpha}_0^H\boldsymbol{\beta}_1 + \boldsymbol{\beta}_0^H\boldsymbol{\alpha}_1\boldsymbol{\alpha}_1^H\boldsymbol{\beta}_1. \end{aligned} \tag{19}$$

The highest eigenvalue of a 2×2 matrix is given by $(T + \sqrt{T^2 - 4D})/2$, where $T = m_{0,0} + m_{1,1}$ and $D = m_{0,0}m_{1,1} - m_{0,1}m_{1,0}$ are the trace and determinant of the matrix, respectively. Therefore, the block coherence $\mu_B(\mathbf{\Phi}_\alpha, \mathbf{\Phi}_\beta)$ is presented as,

$$\begin{aligned} &\mu_B^2\left(\mathbf{\Phi}_\alpha, \mathbf{\Phi}_\beta\right) \\ &= \left(\frac{T + \sqrt{T^2 - 4D}}{2}\right)/4 \\ &= \frac{\left(m_{0,0} + m_{1,1}\right) + \sqrt{\left(m_{0,0} + m_{1,1}\right)^2 - 4\left(m_{0,0}m_{1,1} - m_{0,1}m_{1,0}\right)}}{8} \\ &= \frac{\left(m_{0,0} + m_{1,1}\right) + \sqrt{\left(m_{0,0} - m_{1,1}\right)^2 + 4m_{0,1}m_{1,0}}}{8}. \end{aligned} \tag{20}$$

Therefore, the lower bound of $\mu_{\mathrm{B}}^2(\mathbf{\Phi}_\alpha, \mathbf{\Phi}_\beta)$ can be derived as

$$\mu_{\mathrm{B}}^2(\mathbf{\Phi}_\alpha, \mathbf{\Phi}_\beta) \ge \frac{m_{0,0} + m_{1,1}}{8}. \tag{21}$$

The equality is satisfied if and only if when

$$m_{0,0} = m_{1,1} \tag{22}$$

and

$$\left|m_{0,1}\right| = \left|m_{1,0}\right| = 0. \tag{23}$$

Based on the column rearrangement of $\mathbf{\Phi}_\alpha$ and $\mathbf{\Phi}_\beta$, $\left|\boldsymbol{\alpha}_0^H\boldsymbol{\beta}_0\right|$ and $\left|\boldsymbol{\alpha}_1^H\boldsymbol{\beta}_1\right|$ in (19) are coherences from the identical TS, while $\left|\boldsymbol{\alpha}_0^H\boldsymbol{\beta}_1\right|$ and $\left|\boldsymbol{\alpha}_1^H\boldsymbol{\beta}_0\right|$ are coherences from different TS's, which are also called as *auto-coherence* and *cross-coherence*, respectively. We assume that the auto-coherence is zero to simplify the TS optimization, e.g.,

$$\left|\boldsymbol{\alpha}_0^H\boldsymbol{\beta}_0\right| = \left|\boldsymbol{\alpha}_1^H\boldsymbol{\beta}_1\right| = 0. \tag{24}$$

Then, from (22), we can also derive that the absolute values of different cross-coherence results are the same, i.e.,

$$\left|\boldsymbol{\alpha}_1^H\boldsymbol{\beta}_0\right| = \left|\boldsymbol{\alpha}_0^H\boldsymbol{\beta}_1\right|. \tag{25}$$

According to (24) and (25), the lower bound in (21) can be rewritten as

$$\mu_{\mathrm{B}}(\mathbf{\Phi}_\alpha, \mathbf{\Phi}_\beta) \ge \frac{\left|\boldsymbol{\alpha}_0^H\boldsymbol{\beta}_1\right|}{2}. \tag{26}$$

Therefore, based on (24), (25), and (26), the optimization of μ_B requires a zero value of auto-coherence and a small value of cross-coherence. In our specific case, where the sensing matrix is combined by the transmitted TS's, a realistic objective is to optimize both the auto-coherence and cross-coherence. The two aspects can be achieved independently, and will be investigated in the next two sub-sections.

IV. Auto-Coherence Optimization: TS With Cyclic Extension Structure in an IDFT Pattern

The auto-coherence is minimized in this section. In this context, the TS's for different transmit signals can be designed

and optimized separately to achieve a relatively small value of auto-coherence for each TS. Note that, the sensing matrix based on a specific TS is a Toeplitz matrix according to (10). The IDFT pattern TS with the cyclic extension structure [29] can be adopted in this auto-coherence optimization scheme.

Since we focus on the auto-coherence which is only related to one transmit antenna, the transmit antenna index i is omitted in this section. As a simplified case, when $N = L$, the sensing matrix is a square matrix. If the matrix is a unitary matrix, it is obvious that the auto-coherence of the sensing matrix is zero. However, an arbitrary unitary matrix hardly satisfies the Toeplitz property. Therefore, the TS should be well-designed. Consider a constant amplitude sequence $\boldsymbol{q} = \left[q_0, q_1, ..., q_{L-1}\right]^T$ with normalized amplitude $|q_i| = 1$ for $i = 0, 1, ..., L-1$. The IDFT pattern of TS's adopted in this paper can be denoted as $\boldsymbol{p} = \left[p_0, p_1, ..., p_{L-1}\right]^T$ with elements

$$p_n = \text{IDFT}\,(q_k) = \frac{1}{L}\sum_{k=0}^{L-1} q_k W_L^{-nk}, \tag{27}$$

where $W_L = e^{-j2\pi/L}$ and $0 \le n < L$.

In this condition, the last column of the matrix is an IDFT pattern of TS $\boldsymbol{p}$ with length of L and the rest of the columns of the matrix are all its cyclic extension sequences. Consequently, the sensing matrix can be rewritten as

$$\widetilde{\boldsymbol{\Psi}}|_{N=L} = \begin{bmatrix} p_{L-1} & \cdots & p_2 & p_1 & p_0 \\ p_0 & \cdots & p_3 & p_2 & p_1 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ p_{L-3} & \cdots & p_0 & p_{L-1} & p_{L-2} \\ p_{L-2} & \cdots & p_1 & p_0 & p_{L-1} \end{bmatrix}_{L\times L}. \tag{28}$$

The columns of the matrix are orthogonal, which implies the auto-coherence of the matrix is zero. Under these conditions, the sequence can be represented as

$$\boldsymbol{p}_{\text{opt}}|_{N=L} = \left[p_0, p_1, ..., p_{L-1}, p_0, p_1, ..., p_{L-2}\right]. \tag{29}$$

Under normal circumstances, when $N < L$, it is impossible for all columns to be orthogonal with each other and the auto-coherence can hardly be zero. As a suboptimal approach, the sensing matrix can be truncated through N continuous rows of $\widetilde{\boldsymbol{\Psi}}$. According to the cyclic property of the sensing matrix, the auto-coherence of the new matrix is independent with the starting row to truncate. Without loss of generality, the first N rows of $\widetilde{\boldsymbol{\Psi}}$ are selected. The new sensing matrix and time-domain TS are given in (30) and (31), as shown at the top of the next page. For the time-domain TS, the front part $\left[p_0, p_1, ..., p_{L-1}\right]$ is IDFT of a constant amplitude sequence and the rear part $\left[p_0, p_1, ..., p_{M-L-1}\right]$ is a cyclic extension of the front part.

In this approach, a constant amplitude IDFT pattern of TS's with its cyclic extension is adopted. For each transmit antenna, the TS structure is determined to maintain a relatively small value of auto-coherence. It is worthy mentioning that the auto-coherence cannot be minimized to 0 when $N < L$ in the normal case. The value of auto-coherence is determined by the normalized amplitude sequence. Different initialized sequences may lead to different values of auto-coherence.

Algorithm 1 Genetic Algorithm for TS Optimization.

Inputs:

1) TS length M;
2) Channel length L;
3) Number of transmit antennas N_t;
4) The crossover probability p_c;
5) The mutation probability p_m;
6) Number of individuals in each generation n_d;
7) Maximum number of simulation generations n_g;
8) A discrete probability distribution $\boldsymbol{P}$;

Output:

Generate n_d individuals.
while simulation generation does not reach n_g **do**
  Calculate the fitness value of current generation individuals and sort them descending.
  Distribute the selection probability through $\boldsymbol{P}$ according to the fitness value of the individuals.
  for interaction time from 1 to $n_d/2$ **do**
    Select two individuals based on the selection probability.
    For the chromosome pairs selected, determine whether to be exchanged based on p_c.
    For each chromosomes of an individual, determine whether to perform a mutation based on p_m.
    After crossover and mutation, the two new individuals comprise part of the next generation.
  end for
  Obtain a new generation.
end while
Obtain the smallest fitness value of TS's.

However, an iteration and update process will be performed through a genetic algorithm, which will achieve two goals. One is to determine the suitable TS's with a low value of auto-coherence. The other is to minimize the cross-coherence among the TS's. The process will be presented in the next section.

V. Cross-Coherence Optimization: A Genetic Algorithm

As for the cross-coherence, which reflects the correlation between different TS's, a joint optimization algorithm is needed. To accommodate the requirements of the TS's with low cross-coherence and further lower the values of auto-coherence, a genetic algorithm [36] is applied to optimize the TS's from different transmit antennas. which is addressed in Algorithm **1**.

A. Algorithm Description

The optimization sequences in the algorithm are called individuals. Each individual is the integration of N_t TS's. In detail, N_t TS's with the cyclic extension structure is first randomly generated. And then, the first L elements of each TS with the constant amplitude without IDFT operation are extracted and combined to a new vector. In this way, an

$$\widetilde{\boldsymbol{\Psi}}|_{N=M-L+1} = \begin{bmatrix} p_{L-1} & \cdots & p_{2L-M+1} & p_{2L-M} & p_{2L-M-1} & \cdots & p_1 & p_0 \\ p_0 & \cdots & p_{2L-M+2} & p_{2L-M+1} & p_{2L-M} & \cdots & p_2 & p_1 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{M-L-2} & \cdots & p_0 & p_{L-1} & p_{L-2} & \cdots & p_{M-L} & p_{M-L-1} \\ p_{M-L-1} & \cdots & p_1 & p_0 & p_{L-1} & \cdots & p_{M-L+1} & p_{M-L} \end{bmatrix}_{(M-L+1)\times L} \tag{30}$$

$$\boldsymbol{p}_{\text{opt}}|_{N=M-L+1} = \left[p_0, p_1, ..., p_{L-1}, p_0, p_1, ..., p_{M-L-1}\right] \tag{31}$$

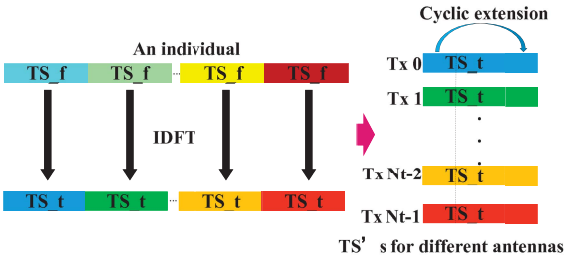

Fig. 3. The relationship between an individual and the TS's from different antennas.

individual is obtained as shown in Fig 3. The individuals can be mapped to a sensing matrix defined in (12). The block coherence of the sensing matrix is the fitness value of the corresponding individual which is the parameter to be optimized.

After n_d individuals are generated, each individual is distributed with a selection probability according to the fitness value and a probability distribution for selection. We randomly select two individuals based on the selection probability, and perform the operations of crossover and mutation for every element of the two individuals one by one. According to the definition of the individuals, the value of individual elements are in the set of $\{1, -1\}$, which is convenient for the operation of crossover and mutation. The operation of crossover is an operation of exchange an element of the two individuals, while the operation of mutation is changing one element of an individual. Appropriate crossover and mutation probabilities are essential to avoid the fitness value going to a local convergence.

The loop ends either when individuals of a new generation are within an error bound or a maximum amount of n_g generations are passed. The convergence of the algorithm can be seen in [36]. After that, an individual with a small block coherence is filtered out, which can be divided into the time-domain TS's of the transmit antennas.

B. Computational Complexity

For the fitness value computation of one individual, block coherence between different blocks is firstly calculated, which has $L(L-1)/2$ block pairs. For each pair of blocks, an operation of matrix multiplication and an $N_t \times N_t$ matrix spectrum norm calculation should be performed, whose computational complexity is $O(N_t^2 N)$. Therefore, the computational complexity for a block coherence is $O((N_t L)^2 N)$. For one generation, n_d fitness values need to be calculated. The sorting, crossover and mutation is not time-consuming process.

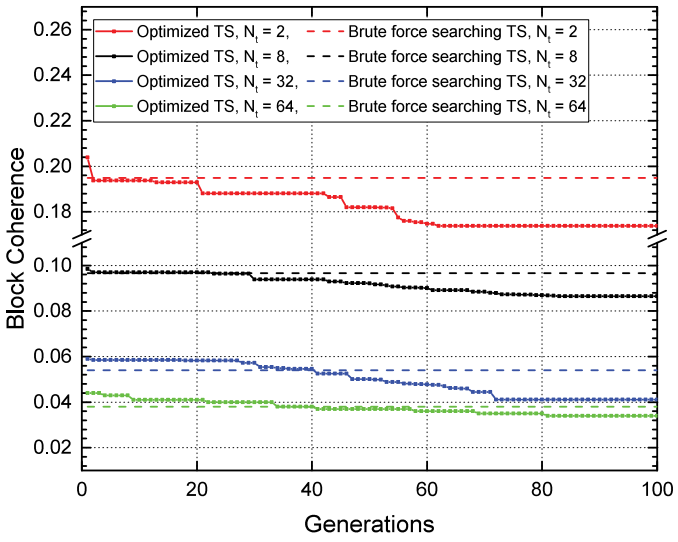

Fig. 4. Coherence convergence property of the proposed GA.

Thus, for n_g generations, the total algorithm computational complexity is $O((N_t L)^2 N n_d n_g)$. It is clear that the algorithm computational complexity is proportional to the times of n_d and n_g. Optimal value selection of the two coefficients can help reduce the complexity and ensure the optimization performance at the same time.

VI. Simulation Results and Discussion

The performance of the proposed TS design and optimization scheme based channel estimation with massive MIMO-OFDM for a mmWave communication system is evaluated in this section. The simulation setup is configured in the uniform linear array-based mm-Wave massive MIMO system where $d = \lambda/2$. In the simulation, the center frequency is 30GHz and the bandwidth is 0.5GHz. The channel model in (2) and (3) are considered with the discrete sparse model in (5). The AoAs/AoDs are assumed to take continuous values, i.e., not quantized, and are uniformly distributed in $[0, 2\pi]$ [33]. The maximum channel delay is $200ns$, which is equivalent to the channel length of 100. The number of paths $s = 3$ [33]. The OFDM data block length is $K = 4096$.

A. Block Coherence of Optimized TS

In Fig. 4, the block coherence convergence process of the proposed genetic algorithm is intuitively illustrated. The channel length of $L = 192$ and PN sequence length of $M = 256$ are simulated. The antenna length of $N_t = 2, 8, 32,$ and 64 are considered, respectively. The

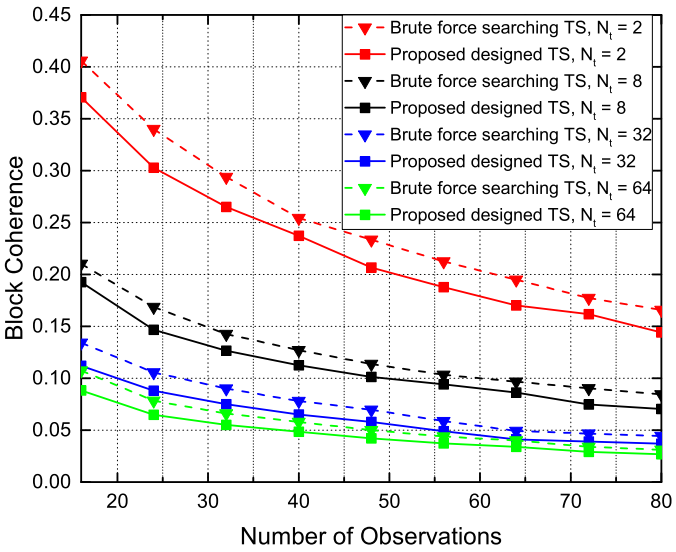

Fig. 5. Optimized block coherence based on the proposed GA.

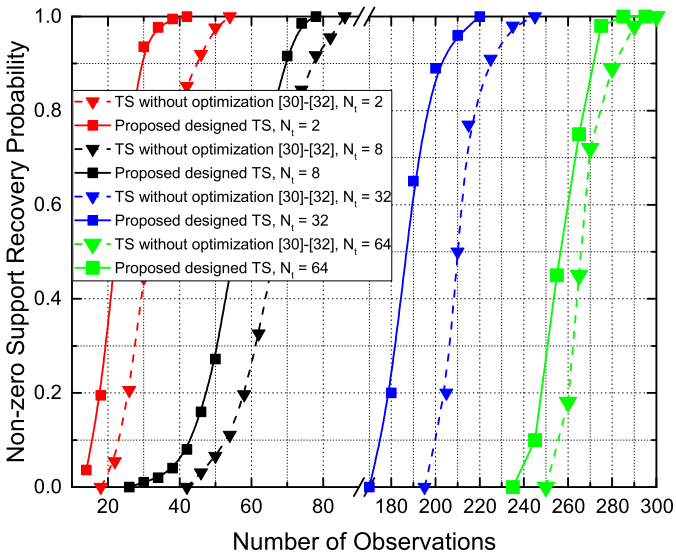

Fig. 6. Recovery probabilities with different number of observations.

optimization performance of brute force searching is also illustrated as a benchmark. It can be seen that the block coherence decreases as the generation number increases. All the examples can come to a convergence after about 70 generation passes. The optimization result by the GA is much lower than that of brute force searching, which proves the effectiveness of the algorithm.

The optimized block coherence of the proposed GA under different simulation parameters is shown in Fig. 5. The minimum values of block coherence using the brute force search method, which costs the same runtime of GA, are also given as comparison. Antenna number of $N_t = 2, 8, 32$, and 64 are simulated, respectively. The number of observations is the length of IBI-free region N. The figure shows that, as the number of observations increase, the sensing matrix of the CS recovery model becomes less flat. Therefore, the value of block coherence become smaller. The figure also shows that the proposed GA can optimize the TS's with a significantly small value of block coherence than those optimized by brute force searching, whatever the number of transmit antennas and observations are. Therefore, the proposed GA is effective in optimizing the block coherence.

B. Performance of Channel Estimation

In the simulation of channel estimation performance, the SCS recovery algorithm is block orthogonal matching pursuit (BOMP) [47], which is an evolutional algorithm to orthogonal matching pursuit (OMP) in CS theory. The columns which are zero or non-zero identically are divided into blocks. The algorithm selects blocks one by one whose columns are the most correlated with the residual vector. The least square (LS) process is performed by the selected columns in blocks and the residual vector can be obtained. After the selection of all the blocks, the LS process is finally performed to acquire the coefficients of the non-zero supports.

Fig. 6 compares the probability of a correct recovery with the optimized TS's under different numbers of transmit antennas. The TS's without optimization [30]–[32] are also

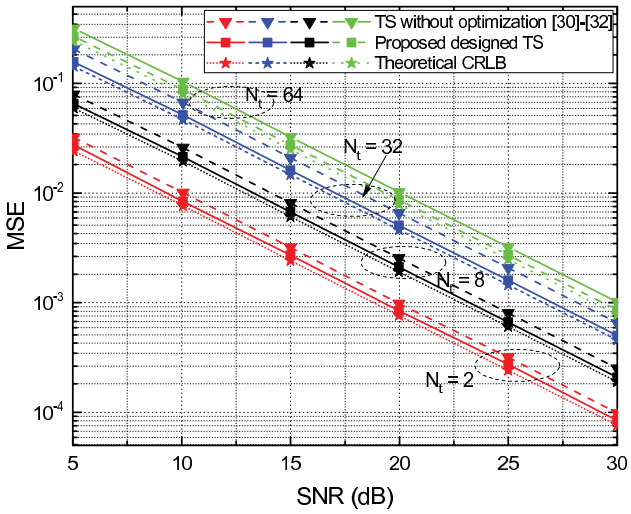

Fig. 7. The MSE of CIR comparison among the TS's under static mmWave channel.

illustrated as the benchmark. The correct recovery here is defined as the accurate estimation of the non-zero support. The signal-to-noise radio (SNR) is configured as 25 dB. Different numbers of observations are simulated. It can be seen that all the TS's under different numbers of transmit antennas outperform the TS's without optimization. The performance improvement mainly comes from low coherence between different blocks of sensing matrix, which facilitates more accurate block sparse channel recovery. In the concrete, the proposed optimized TS's reach a successful recovery probability of 0.9 at the observation number of less than 30, 70, 200, and 270, for the transmitted antenna number of $N_t = 2, 8, 32$, and 64, respectively. Furthermore, TS's without optimization require at least 18, 10, 15, and 11 more observations for the same cases. Therefore, the results indicate that the proposed designed and optimized TS's are more efficient in block sparse signal recovery than those without optimization.

Fig. 7 presents the MSE of CIR for different TS patterns under a typical static mmWave channel. The performances

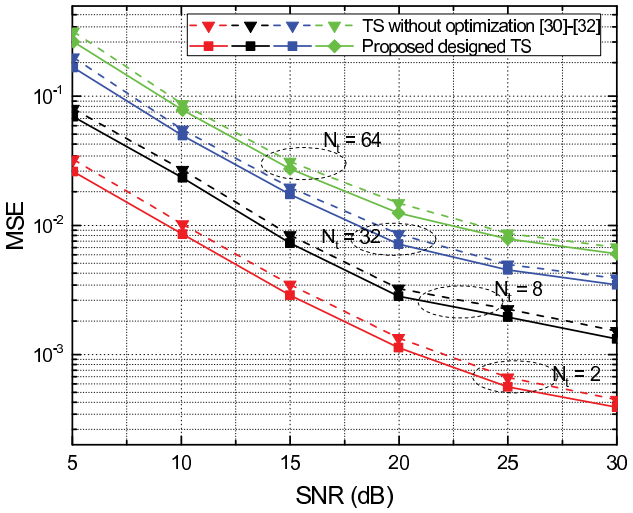

Fig. 8. The MSE of CIR comparison among the TS's under time-variant mmWave channel.

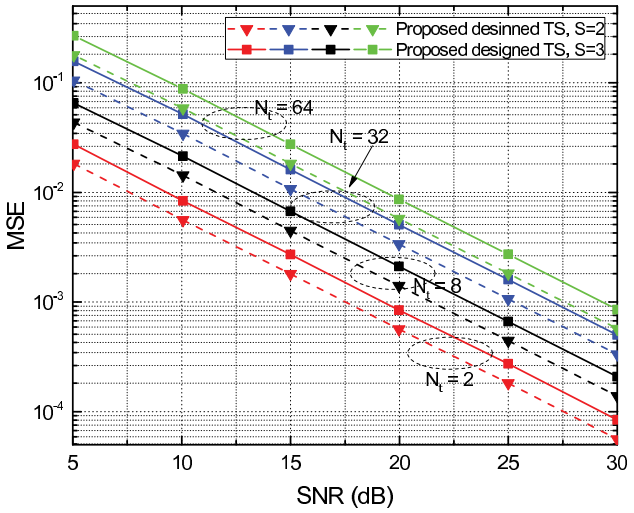

Fig. 9. The MSE performance with different values of channel sparsity.

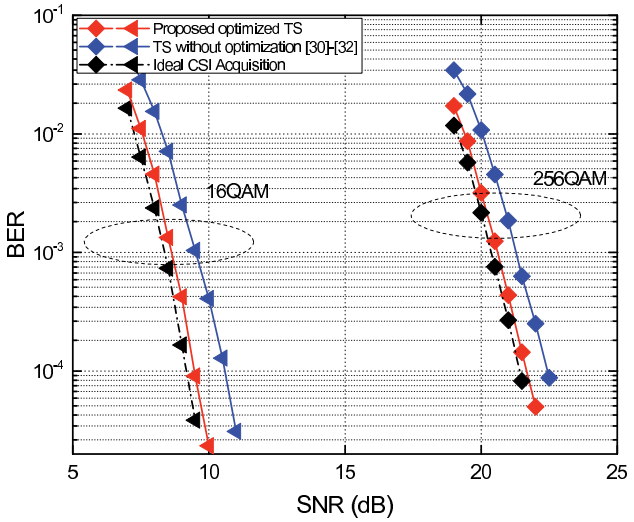

Fig. 10. BER performance comparison under static channel.

of TS's without optimization [30]–[32] are shown for comparison, while the ideal CRLB's are also illustrated as the benchmarks (CRLB $= SN_t\sigma^2/N$, whose proof can be seen in Appendix). The length of the TS is $M = 256, 384, 512,$ and 768 for $N_t = 2, 8, 32,$ and 64, respectively. As the antenna number becomes large, the MSE performance have a little degradation naturally. However, the channel estimation performance by utilizing the proposed design and optimization TS can always approach the CRLB. In detail, tt can be seen that the proposed TS's have over 1 dB improvement compared with the TS's without optimization when the MSE of 10^{-2} is considered. Moreover, the MSE is less than 0.2 dB away from the theoretical CRLB above 25 dB SNR. Therefore, the proposed TS design and optimization scheme can have a superior performance in channel estimation for massive MIMO-OFDM systems.

Fig. 8 presents the MSE of CIR for different TS patterns under the time-variant mmWave channel. The performances of TS's without optimization [30]–[32] are shown for comparison. The length of the TS is the same with those in Fig. 7 corresponding to different number of antennas. It can be seen that, under the time-variant channel, the proposed TS's also have about 1 dB improvement compared with the TS's without optimization when the MSE of 10^{-2} is considered. The difference between the simulations of the two channels is that, the MSE performance degrades in high SNR levels under the time-variant channel. This is mainly because that, the variant channel in time domain can reflect as noise when performing channel estimation. Therefore, in high level of SNR, the MSE results will no longer decrease linearly and the performance can hardly be as good as that under static channel. After all, the optimized TS's can outperform the ones without optimization in channel estimation under time-variant mmWave channel.

Fig. 9 shows the MSE performance with different values of channel sparsity. The sparsity level $S = 2$ and $S = 3$ are simulated for comparison. In the simulation result, the MSE is lower when the sparsity level is small. About 2 dB gain can be obtained when $S = 2$ is utilized compared with that when $S = 3$, which means the channel estimation performs better and the accuracy is higher when the sparsity is smaller.

Fig. 10 compares the BER performance when low density parity check code (LDPC) with code rate of 0.4 and code length of 7493 specified by DTMB are adopted [37] in a 64×64 MIMO system with 256QAM and 16QAM constellations under the static channel. The BER performances with the ideal CSI are also illustrated as benchmarks. The ideal CSI means the channel estimation result when knowing the accurate non-zero support. It can be observed that the CS-based channel estimation scheme can support high-order modulation schemes like 256QAM well. Moreover, for low-order modulation schemes like 16QAM, the CS-based channel estimation scheme has great performance with low SNR requirement. The proposed TS design schemes outperform the conventional non-optimization ones [30]–[32], which have about 1.3 dB and 1 dB gain when the BER of 10^{-3} is considered for 16QAM and 256QAM, respectively. Furthermore, the two schemes are both only about 0.2 dB away from the ideal cases.

Fig. 11 illustrates the BER comparison under the time-variant channel. LDPC is also utilized with the same code

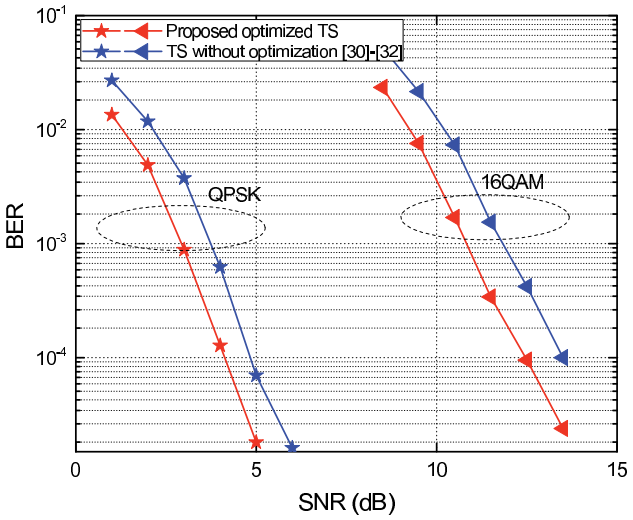

Fig. 11. BER performance comparison under time-variant channel.

rate and code length in Fig. 10 for a 64×64 MIMO system. QPSK and 16QAM constellations are considered. It can be observed that the proposed TS design schemes outperform the conventional non-optimization ones, which have about 0.6 dB and 1 dB gain at the BER of 10^{-3} for QPSK and 16QAM schemes, respectively. This means our method for TS design and optimization is also available for time-variant mmWave channel estimation in massive MIMO-OFDM system.

VII. Conclusions

In this paper, the TS design and optimization algorithm suitable for massive MIMO-OFDM channel estimation in mmWave communications is proposed. Considering the spatial correlation between different channels, the spatial correlated channel model is utilized and the SCS method is adopted by utilizing the IBI-free region to perform the channel estimation. In order to improve the recovery accuracy, the auto-coherence and cross-coherence of the blocks are proposed as two key merit factors, which is a new perspective to optimize the block coherence of sensing matrix. The application of the proposed IDFT sequence and the genetic algorithm are two main approaches to optimize the merit factors. Simulation results indicate that the proposed TS design and optimization method can significantly reduce the block coherence with different numbers of transmit antennas. The optimized value of block coherence is rather smaller than that of brute force searching TS. Furthermore, by utilizing the optimized TS's in different transmit antennas, the channel estimation can have better performance in terms of correct recovery probability, MSE, and BER, which means smaller length of TS is required and higher capacity of communication can be achieve. Better performance can be realized under both static channel and the time-variant channel. Hence, the proposed scheme is a promising physical layer technique for the mmWave communications which can be applied in the future 5G communications to improve the system performance. Some of the promising research is detailed below which outline some important areas requiring some future work.

A. Multi-User MIMO (MU-MIMO) System

The proposed time-domain training sequence design based channel estimation scheme is applied to single-user MIMO (SU-MIMO) scenarios. Fortunately, the SU-MIMO scenarios proposed in this paper can be directly extended to MU-MIMO condition. To make full use of the characteristic of multi-users, the relationship among different users should also be considered, as well as their joint influence to the whole system.

B. Pilot Contamination in Cellular Network

The frequency-domain pilot contamination [48] does not exist in our proposed scheme due to the utilization of time-domain training sequence. Actually, the contamination of time-domain training sequence will also appear in the multi-cell scenarios. Similar approach with frequency-domain pilot contamination elimination can be adopted, such as, using different time-blocks for different cells [48]. The study of pilot contamination will be an important research direction in the future.

C. Joint Design for Multiple Antennas

In this paper, the reception at different receive antennas are considered independently while a joint design is not investigated. Moreover, the training sequence design at transmitter side for beamforming will also be considered in future work.. Further research can be performed for the training sequence joint design for the multiple antennas.

Appendix
Proof of CRLB of MSE for Static Channel

In the ideal case, the non-zero support is estimated correctly, which can be denoted as $D(\|D\|_0 = SN_t)$. Therefore, the entries outside the support D are set to zeros. Ignoring the noise $\boldsymbol{n}$, equation (12) is simplified as

$$\boldsymbol{r} = \boldsymbol{\Phi}_D \boldsymbol{g}_D, \tag{1}$$

which can be estimated by solving an over-determined equation under the Maximum Likelihood (ML) criterion.

$$\boldsymbol{g}_{\text{est}} = \boldsymbol{\Phi}_D^{+}\mathbf{r} = (\boldsymbol{\Phi}_D^H \boldsymbol{\Phi}_D)^{-1}\boldsymbol{\Phi}_D \boldsymbol{r}. \tag{2}$$

Then, the CRLB of $\boldsymbol{g}$ can be denoted as

$$\text{CRLB} = \text{E}\{\|\boldsymbol{g}_{\text{est}} - \boldsymbol{g}_D\|_2\} = \frac{SN_t}{N}\sigma^2. \tag{3}$$

References

[1] A. Osseiran *et al.*, "Scenarios for 5G mobile and wireless communications: The vision of the METIS project," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 26–35, May 2014.

[2] C. Park and T. S. Rappaport, "Short-range wireless communications for next-generation networks: UWB, 60 GHz millimeter-wave WPAN, and ZigBee," *IEEE Wireless Commun.*, vol. 14, no. 4, pp. 70–78, Aug. 2007.

[3] T. S. Rappaport, R. W. Heath, Jr., R. C. Daniels, and N. Murdock, *Millimeter Wave Wireless Communications*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2014.

[4] X. Gao, L. Dai, S. Han, C. L. I, and R. W. Heath, Jr., "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, Apr. 2016.

[5] M. K. Samimi and T. S. Rappaport, "3-D millimeter-wave statistical channel model for 5G wireless system design," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 7, pp. 2207–2225, Jul. 2016.

[6] A. I. Sulyman, A. Alwarafy, G. R. MacCartney, T. S. Rappaport, and A. Alsanie, "Directional radio propagation path loss models for millimeter-wave wireless networks in the 28-, 60-, and 73-GHz bands," *IEEE Trans. Wireless Commun.*, vol. 15, no. 10, pp. 6939–6947, Oct. 2016.

[7] D. S. Palguna, D. J. Love, T. A. Thomas, and A. Ghosh, "Millimeter wave receiver design using low precision quantization and parallel $\Delta\Sigma$ architecture," *IEEE Trans. Wireless Commun.*, vol. 15, no. 10, pp. 6556–6569, Oct. 2016.

[8] E. Perahia, C. Cordeiro, M. Park, and L. L. Yang, "IEEE 802.11ad: Defining the next generation multi-Gbps Wi-Fi," in *Proc. 7th IEEE Consum. Commun. Netw. Conf.*, Las Vegas, NV, USA, Jan. 2010, pp. 1–5.

[9] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[10] D. T. Emerson, "The work of Jagadis Chandra Bose: 100 years of mm-wave research," in *IEEE MTT-S Int. Microw. Symp. Dig.*, vol. 2. Denver, CO, USA, Jun. 1997, pp. 553–556.

[11] J. H. van Vleck, "The absorption of microwaves by oxygen," *Phys. Rev.*, vol. 71, no. 7, pp. 413–424, Apr. 1947.

[12] A. R. Tharek and J. P. McGeehan, "Outdoor propagation measurements in the millimetre wave band at 60 GHz," in *Proc. 6th Conf. Military Microw.*, London, U.K., Jul. 1988, pp. 43–48.

[13] S. W. Wales and D. C. Rickard, "Wideband propagation measurements of short range millimetric radio channels," *Electron. Commun. Eng. J.*, vol. 5, no. 4, pp. 249–254, Aug. 1993.

[14] "Regulations for enforcement of the radio law 6-4-2 specified low power radio station, (11) 59–66 band," Ministry Public Manage., Home Affairs, Posts Telecommun., Japan, Tech. Rep., 2000.

[15] X. Gao, L. Dai, C. Yuen, and Z. Wang, "Turbo-like beamforming based on tabu search algorithm for millimeter-wave massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5731–5737, Jul. 2016.

[16] M. Peng, Y. Li, Z. Zhao, and C. Wang, "System architecture and key technologies for 5G heterogeneous cloud radio access networks," *IEEE Netw.*, vol. 29, no. 2, pp. 6–14, Mar./Apr. 2015.

[17] Y. J. Hong, J. Kim, and D. K. Sung, "Two-dimensional channel estimation and prediction for scheduling in cellular networks," *IEEE Trans. Veh. Technol.*, vol. 58, no. 6, pp. 2727–2740, Jul. 2009.

[18] Y. Wang, C. Li, Y. Huang, D. Wang, T. Ban, and L. Yang, "Energy-efficient optimization for downlink massive MIMO FDD systems with transmit-side channel correlation," *IEEE Trans. Veh. Technol.*, vol. 65, no. 9, pp. 7228–7243, Sep. 2016.

[19] W. Liu, S. Jin, C.-K. Wen, M. Matthaiou, and X. You, "A tractable approach to uplink spectral efficiency of two-tier massive MIMO cellular HetNets," *IEEE Commun. Lett.*, vol. 20, no. 2, pp. 348–351, Feb. 2016.

[20] L. Dai, Z. Wang, and Z. Yang, "Spectrally efficient time-frequency training OFDM for mobile large-scale MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 251–263, Feb. 2013.

[21] 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.

[22] C. Qi, L. Wu, Y. Huang, and A. Nallanathan, "Joint design of pilot power and pilot pattern for sparse cognitive radio systems," *IEEE Trans. Veh. Technol.*, vol. 64, no. 11, pp. 5384–5390, Nov. 2015.

[23] J. Meng, W. Yin, Y. Li, N. T. Nguyen, and Z. Han, "Compressive sensing based high-resolution channel estimation for OFDM system," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 1, pp. 15–25, Feb. 2012.

[24] Z. Han, H. Li, and W. Yin, *Compressive Sensing for Wireless Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2013.

[25] W. Ding *et al.*, "Spectrally efficient CSI acquisition for power line communications: A Bayesian compressive sensing perspective," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 7, pp. 2022–2032, Jul. 2016.

[26] S. Liu, F. Yang, W. Ding, and J. Song, "Double kill: Compressive-sensing-based narrow-band interference and impulsive noise mitigation for vehicular communications," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5099–5109, Jul. 2016.

[27] S. Liu, F. Yang, and J. Song, "Narrowband interference cancelation based on priori aided compressive sensing for DTMB systems," *IEEE Trans. Broadcast.*, vol. 61, no. 1, pp. 66–74, Mar. 2015.

[28] F. Yang, J. Gao, and S. Liu, "Priori aided compressed sensing-based clipping noise cancellation for ACO-OFDM systems," *IEEE Photon. Technol. Lett.*, vol. 28, no. 19, pp. 2082–2085, Oct. 1, 2016.

[29] X. Ma, F. Yang, W. Ding, and J. Song, "Novel approach to design time-domain training sequence for accurate sparse channel estimation," *IEEE Trans. Broadcast.*, vol. 62, no. 3, pp. 512–520, Sep. 2016.

[30] W. Ding, F. Yang, C. Pan, L. Dai, and J. Song, "Compressive sensing based channel estimation for OFDM systems under long delay channels," *IEEE Trans. Broadcast.*, vol. 60, no. 2, pp. 313–321, Jun. 2014.

[31] W. Ding, F. Yang, W. Dai, and J. Song, "Time–frequency joint sparse channel estimation for MIMO-OFDM systems," *IEEE Commun. Lett.*, vol. 19, no. 1, pp. 58–61, Jan. 2015.

[32] W. Ding, F. Yang, S. Liu, X. Wang, and J. Song, "Nonorthogonal time–frequency training-sequence-based CSI acquisition for MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5714–5719, Jul. 2016.

[33] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[34] Z. Gao, C. Hu, L. Dai, and Z. Wang, "Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1259–1262, Jun. 2016.

[35] M. Kokshoorn, H. Chen, P. Wang, Y. Li, and B. Vucetic, "Millimeter wave MIMO channel estimation using overlapped beam patterns and rate adaptation," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 601–616, Feb. 2017.

[36] J. E. Beasley and P. C. Chu, "A genetic algorithm for the set covering problem," *Eur. J. Oper. Res.*, vol. 94, no. 2, pp. 392–404, Oct. 1996.

[37] J. Song *et al.*, "Technical review on Chinese digital terrestrial television broadcasting standard and measurements on some working modes," *IEEE Trans. Broadcast.*, vol. 53, no. 1, pp. 1–7, Mar. 2007.

[38] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Multilevel millimeter wave beamforming for wireless backhaul," in *Proc. IEEE GLOBECOM Workshops (GC Wkshps)*, Houston, TX, USA, Dec. 2011, pp. 253–257.

[39] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4053–4085, Sep. 2011.

[40] M. Masood, L. H. Afify, and T. Y. Al-Naffouri, "Efficient coordinated recovery of sparse channels in massive MIMO," *IEEE Trans. Signal Process.*, vol. 63, no. 1, pp. 104–118, Jan. 2015.

[41] Y. Barbotin, A. Hormati, S. Rangan, and M. Vetterli, "Estimation of sparse MIMO channels with common support," *IEEE Trans. Commun.*, vol. 60, no. 12, pp. 3705–3716, Dec. 2011.

[42] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[43] Z. Ben-Haim, Y. C. Eldar, and M. Elad, "Coherence-based performance guarantees for estimating a sparse vector under random noise," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5030–5043, Oct. 2010.

[44] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4680–4688, Jul. 2011.

[45] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[46] Y. C. Eldar and H. Bolcskei, "Block-sparsity: Coherence and efficient recovery," in *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 2885–2888.

[47] W. Hou and C. W. Lim, "Structured compressive channel estimation for large-scale MISO-OFDM systems," *IEEE Commun. Lett.*, vol. 18, no. 5, pp. 765–768, May 2014.

[48] E. Björnson, E. G. Larsson, and M. Debbah, "Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?" *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1293–1308, Feb. 2016.

Xu Ma (S'16) received the B.S.E. degree from the Department of Electronic Engineering, Tsinghua University, Beijing, China, in 2014, where he is currently pursuing the Ph.D. degree with the DTV Technology Research and Development Center. His research interests lie in the field of the power line communication, visible light communication, and sparse signal processing theory in wireless communications.

Fang Yang (M'11–SM'13) received the B.S.E. and Ph.D. degrees in electronic engineering from Tsinghua University, Beijing, China, in 2005 and 2009, respectively. He is currently an Associate Professor with the Research Institute of Information Technology, Tsinghua University. He has authored over 120 peer-reviewed journal and conference papers. He holds over 40 Chinese patents and two PCT patents. His research interests lie in the fields of channel coding, channel estimation, interference cancelation, and signal processing techniques for communication system, especially in power line communication, visible light communication, and digital television terrestrial broadcasting. He received the IEEE Scott Helt Memorial Award (Best Paper Award in IEEE TRANSACTIONS IN BROADCASTING) in 2015. He is the Secretary General of Sub-Committee 25 of the China National Information Technology Standardization (SAC/TC28/SC25).

Sicong Liu (S'15) received the B.S.E. degree in electronic engineering from the Department of Electronic Engineering, Tsinghua University, Beijing, China, where he is currently pursuing the Ph.D. degree in electronic engineering with the DTV Technology Research and Development Center, Department of Electronic Engineering. His research interests lie in 5G wireless technologies, sparse signal processing and its applications on power line communications, smart grid communications, and noise and interference mitigation.

Jian Song (M'06–SM'10–F'16) received the B.Eng. and Ph.D. degrees in electrical engineering from Tsinghua University, Beijing China, in 1990 and 1995, respectively. He was with The Chinese University of Hong Kong, Hong Kong, and with the University of Waterloo, Canada, in 1996 and 1997, respectively. He was with Hughes Network Systems, USA, for seven years before joining the faculty team in Tsinghua University, in 2005, as a Professor. He is currently the Director of the Tsinghua's DTV Technology Research and Development Center. He has been involved in the quite different areas of fiber-optic, satellite and wireless communications, and the power line communications. He has authored or co-authored over 200 peer-reviewed journal and conference papers. He holds two U.S. and over 40 Chinese patents. His current research interest is in the area of digital TV broadcasting. He is a Fellow of IET.

Zhu Han (S'01–M'04–SM'09–F'14) received the B.S. degree in electronic engineering from Tsinghua University, in 1997, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Maryland, College Park, MD, USA, in 1999 and 2003, respectively.

From 2000 to 2002, he was a Research and Development Engineer of JDSU, Germantown, MD, USA. From 2003 to 2006, he was a Research Associate with the University of Maryland. From 2006 to 2008, he was an Assistant Professor with Boise State University, ID, USA. He is currently a Professor with the Electrical and Computer Engineering Department and with the Computer Science Department, University of Houston, Houston, TX, USA. His current research interests include wireless resource allocation and management, wireless communications and networking, game theory, big data analysis, security, and smart grid. He was a recipient of the NSF Career Award in 2010, the Fred W. Ellersick Prize from the IEEE Communication Society in 2011, the EURASIP Best Paper Award for the *Journal on Advances in Signal Processing* in 2015, the IEEE Leonard G. Abraham Prize in the field of Communications Systems (Best Paper Award in IEEE JSAC) in 2016, and several best paper awards in IEEE conferences. He is currently an IEEE Communications Society Distinguished Lecturer.