

# Robust Gridless Estimation of Angles and Delays for Full-Dimensional Wideband mmWave Channels

TING-MING YANG <sup>id</sup> (Senior Member, IEEE), AND YUAN-PEI LIN <sup>id</sup> (Senior Member, IEEE)

Department Electrical Engineering, National Yang Ming Chiao Tung University, Hsinchu 30010, Taiwan

CORRESPONDING AUTHOR: YUAN-PEI LIN (email: he02515939@gmail.com).

This work was supported by National Science and Technology Council, Taiwan, under Grant 109-2221-E-009-104-MY3.

**ABSTRACT** In this paper, we consider robust channel estimation for a millimeter wave (mmWave) massive MIMO system with uniform planar arrays (UPA). For many gridless angle estimation methods of mmWave channels, the channel gains needs to be time-invariant during training. We propose a gridless method that is applicable to time-invariant and time-varying channels, and the proposed method is robust to channel variations. This is done by employing time-varying precoders and combiners that are banded block Toeplitz (BBT) matrices. The advantage of time-varying BBT precoders is two-fold. First, with time-varying BBT precoding we can fabricate time-varying channels. Thus the channel can be estimated as if it is a time-varying channel whether the actual channel is time-varying or not. We will show that, thanks to fabricated time variation in the channel, a small training overhead can be used. Secondly, BBT precoding and combining preserve the shift invariance property of the received data matrix. Due to the preservation of shift invariance property, we can employ most existing direction finding methods for estimating the angles and delays, including low complexity gridless subspace based algorithms. Simulation results are given to show that the proposed time-varying BBT precoding provides robust and accurate channel estimates with a low complexity.

**INDEX TERMS** MIMO communication, Massive MIMO, Millimeter wave communication.

## I. INTRODUCTION

High data rates have been achieved in mmWave communication systems [1] by employing a large number of antennas for beamforming. Due to heavy path loss, an mmWave channel typically consists of a few dominant paths. The estimation of an mmWave channel becomes a problem of finding the angles and gains associated with the paths. Experiments have shown that the angles of mmWave channels experience large-scale fading whereas the path gains experience small-scale fading [2]. In other words, AoA and AoD are relatively stationary, but the path gains can vary faster, leading to a time-varying channel.

In the design of angle estimation algorithms, there are often some implicit assumptions on the channel. For example, direction finding of narrowband mmWave systems has been developed for time-invariant channels [3], [4], [5], [6], [7], [8], [9] and time-varying channels [10], [11]. For time-invariant channels, angles estimation methods have been developed

for uniform linear array (ULA) and uniform planar array (UPA). The sparse nature of mmWave has been exploited in compressed sensing based methods for angle estimation [3], [4]. In these methods, the angles are approximated by values drawn from dictionaries or grids, which restrict the estimation resolution. To eliminate the limitation of grids in compressed sensing based methods, super resolution methods have been proposed in [5], [6] based on atomic norm minimization and in [7] based on an iterative reweighted method. Super resolution angle estimation is proposed in [8] by designing the precoder and combiner so that efficient subspace-based methods can be applied for angles estimation. Beamspace angle estimation for multi-user transmission is considered in [9] when the transmitter is equipped with UPA and receiver with a single antenna. For time-varying channels, the path gains vary during training. In [10], [11] the path gains in different training blocks are assumed to be uncorrelated. When the path gains are uncorrelated, the covariance of the received

data has a Toeplitz structure, which can be utilized for noise reduction [10]. Beamspace approaches to direction finding for time-varying mmWave channels are considered in [11]. The assumption of uncorrelated path gains requires the training time be longer than the channel coherence time so that there is enough variation in the path gains during training [10], [11]. The estimation accuracy is affected when the path gains are almost static due to little user movement. For methods that are developed specifically for time-varying channels or time-invariant channels, the estimation result is sensitive to the time-variation of the channel. Direction finding methods that are not sensitive to the time-variation of the channel have been considered in [12], [13]. A hierarchical beam training method using successive beams of different beamwidth is considered in [12]. The estimation accuracy is limited by the underlying grids. In [13] the precoder is obtained by selecting the columns of a Toeplitz matrix according to the spacing of a sparse linear array. An enlarged virtual ULA can be constructed for high resolution angle estimation. However, a large number of training vectors is generally needed therein as the estimation is based on statistics, and the method does not generalize to wideband channels or when the array is UPA. For wideband transmission, the delays of different paths need to be estimated as well. Subspace based and compressed sensing based channel estimation methods have been proposed for either time-invariant channels and time-varying channels. Subspace based methods are considered in [14], [15] for time invariant channels using closed-loop two-stage estimation. With the aid of feedback after the first stage, better estimates can be obtained. Subspace based methods have also been used for time-varying wideband channels and the variation in path gains is utilized for joint estimation of angles and delays [16], [17]. Compressive sensing based methods for MIMO-OFDM systems are proposed in [18], [19] for time-invariant systems and in [20], [21] for time-varying systems. Two channel estimation methods are developed in [18], one using a time domain approach that performs a 3D grid search, and the other a frequency domain approach that uses orthogonal matching pursuit (OMP) for each OFDM subcarrier. The estimation accuracy of the frequency domain approach can be further improved by exploiting the fact that AoA and AoD are the same in all subcarriers [19]. To track time-varying path gains, LMS adaptation is employed in [20] and Bayesian learning is used in [21]. Due to adaptation, the estimation result is not sensitive to the channel variations. However the updates performed in these adaptive methods demand a high complexity as computations of large matrices are involved. In addition, in these compressed sensing based works [18], [19], [20], [21], path delays can not be estimated and the information is embedded in the subcarrier channels of the MIMO-OFDM system. In this case, the path gains of all the subcarrier channels need to be fed back to the transmitter and a large feedback overhead is needed [14].

In this paper, we propose a robust angle estimation method that is applicable to time-invariant and time-varying channels for mmWave systems equipped with UPA. We use precoders

and combiners that are banded block Toeplitz matrices (BBT). We show that BBT precoding allows us to have partial control of the equivalent path gains. The equivalent channel is always time varying by employing time-varying BBT precoding. As a result the estimation of angles for mmWave channels is converted to a classic angle estimation problem that can be solved by many useful methods in the literature. Exploiting the time-varying property of the equivalent channel, the angles can be estimated using a small overhead, even if the channel is time-invariant. For wideband transmission, angles as well as delays need to be estimated. We show that we can design BBT precoders so that an artificial time-varying channel can be fabricated for wideband channels too. The path delays can be estimated along with the angles and they can be fed back with a small feedback overhead. The contributions of this paper are summarized as follows:

- We propose to use time-varying BBT precoding so that both time-varying and time-invariant mmWave channels can be reliably estimated. As no assumption of prior channel information is used, it is suitable for angle estimation in the initial stage of channel acquisition.
- Through BBT precoding, the estimation of angles and delays of mmWave channels is converted to a classic harmonic retrieval problem that can be solved using well-developed methods in the literature with a low complexity. Adaptive estimation methods that are applicable to both time-invariant and time-varying wideband mmWave channels have also been developed in [20], [21]. However the algorithms in [20], [21] require the computations of large matrices and a high complexity is needed.
- Time-varying BBT precoding leads to artificial time-varying path gains and the proposed method needs only a small number of training blocks. For example, for narrowband channels, the minimum number of training blocks is equal to the number of paths  $L$ , for either time-invariant and time-varying channels.
- The path delays are estimated jointly with the angles. As the delays are explicitly estimated, they can be fed back to the transmitter directly using a small overhead. When the path delays are not estimated [18], [19], [20], [21], but embedded in subcarrier channels, a large feedback overhead is required [14].

*Notation:* The variance of a random variable  $x$  is denoted as  $\sigma_x^2$ . The 2-norm of a vector  $\mathbf{f}$  is denoted as  $\|\mathbf{f}\|$  and the  $k$ -th entry of  $\mathbf{f}$  is  $[\mathbf{f}]_k$ . The notation  $\mathbf{A}^T$ ,  $\mathbf{A}^*$ ,  $\mathbf{A}^H$  denote, respectively, the transpose, the conjugate, the conjugate transpose.  $\mathbf{I}_n$  denotes an  $n \times n$  identity matrix. Given an  $m \times n$  matrix  $\mathbf{X}$ ,  $\text{vec}(\mathbf{X})$  is the  $nm \times 1$  vectorized version of  $\mathbf{X}$ . The notation  $\mathbf{A} \odot \mathbf{B}$  denotes the Khatri-Rao product and  $\mathbf{A} \otimes \mathbf{B}$  the Kronecker product of two matrices  $\mathbf{A}$  and  $\mathbf{B}$ . The expectation of a random variable is denoted by  $E[\cdot]$ . A useful property concerning the two products is as follows. Let  $\mathbf{X} = \mathbf{ABC}$ , then  $\text{vec}(\mathbf{X}) = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}(\mathbf{B})$ . When  $\mathbf{B}$  is a diagonal matrix, we have  $\text{vec}(\mathbf{X}) = (\mathbf{C}^T \odot \mathbf{A})\mathbf{b}$ , where  $\mathbf{b}$  is a column vector that consists of the diagonal elements of  $\mathbf{B}$ .

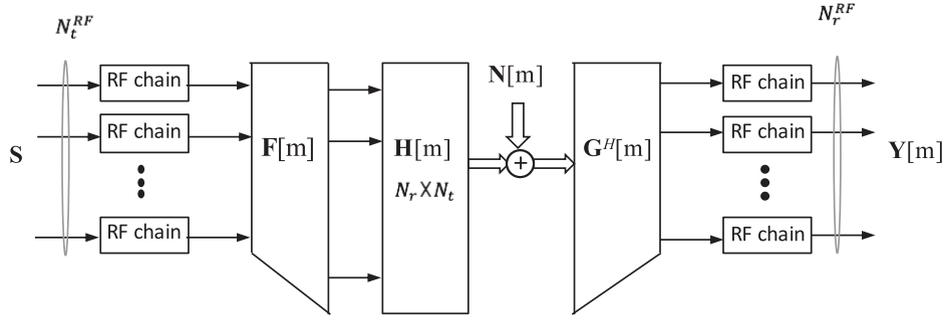


FIGURE 1. A MIMO communication system with hybrid precoding.

*Outline of the paper:* A system model for narrowband transmission is given in Section II. In Section III, we introduce BBT precoding and develop angle estimation for narrowband channels. Section IV addresses the estimation of angles and path delays for wideband channels. Simulation results are given in Section V and a conclusion in Section VI.

## II. SYSTEM MODEL FOR NARROWBAND TRANSMISSION

Consider the narrowband wireless system with  $N_t$  transmit antennas and  $N_r$  receive antennas in Fig. 1. For the  $m$ -th training block, the channel is modeled by an  $N_r \times N_t$  matrix  $\mathbf{H}[m]$ . We adopt the geometric channel representation that is useful for modeling mmWave propagation of  $L$  paths [2],

$$\mathbf{H}[m] = \mathbf{A}_r \mathbf{D}_\alpha[m] \mathbf{A}_t^H, \quad (1)$$

where  $\mathbf{D}_\alpha[m]$  is an  $L \times L$  diagonal matrix whose  $i$ -th diagonal element  $[\mathbf{D}_\alpha[m]]_{i,i} = \alpha_i[m]$  is the complex gain of the  $i$ -th path. For the case the transmitter and receiver are each equipped with a uniform linear array (ULA), the two matrices  $\mathbf{A}_r$ , of size  $N_r \times L$ , and  $\mathbf{A}_t$ , of size  $N_t \times L$ , are given by

$$\begin{aligned} \mathbf{A}_r &= \begin{pmatrix} \mathbf{a}_{N_r}(u_1) & \mathbf{a}_{N_r}(u_2) & \cdots & \mathbf{a}_{N_r}(u_L) \end{pmatrix}, \\ \mathbf{A}_t &= \begin{pmatrix} \mathbf{a}_{N_t}(p_1) & \mathbf{a}_{N_t}(p_2) & \cdots & \mathbf{a}_{N_t}(p_L) \end{pmatrix}, \end{aligned} \quad (2)$$

where

$$\mathbf{a}_{N_t}(x) = \begin{pmatrix} 1 & e^{jx} & \cdots & e^{j(N-1)x} \end{pmatrix}^T. \quad (3)$$

The angle parameters  $u_i$  can be expressed in terms of AoA  $\theta_i^t$  and the parameters  $p_i$  in terms of AoD  $\theta_i^t$ ,

$$u_i = (2\pi d/\lambda) \cos \theta_i^t, \quad p_i = (2\pi d/\lambda) \cos \theta_i^t,$$

where  $d$  is the antenna spacing and  $\lambda$  is the wavelength. On the other hand, consider the case when the transmitter and receiver are each equipped with a uniform planar array (UPA) in the  $yz$ -plane [14]. Suppose the UPA at the receiver is  $N_{ry} \times N_{rz}$  ( $N_{ry}$  in the  $y$  direction and  $N_{rz}$  in the  $z$  direction) and  $N_r = N_{ry}N_{rz}$ , then  $\mathbf{A}_r$ , of size  $N_r \times L$ , is given by

$$\mathbf{A}_r = \mathbf{A}_{ry} \odot \mathbf{A}_{rz}, \quad (4)$$

where  $\mathbf{A}_{ry}$ , of size  $N_{ry} \times L$ , and  $\mathbf{A}_{rz}$ , of size  $N_{rz} \times L$ , are defined as

$$\begin{aligned} \mathbf{A}_{ry} &= \begin{pmatrix} \mathbf{a}_{N_{ry}}(u_1^y) & \mathbf{a}_{N_{ry}}(u_2^y) & \cdots & \mathbf{a}_{N_{ry}}(u_L^y) \end{pmatrix}, \\ \mathbf{A}_{rz} &= \begin{pmatrix} \mathbf{a}_{N_{rz}}(u_1^z) & \mathbf{a}_{N_{rz}}(u_2^z) & \cdots & \mathbf{a}_{N_{rz}}(u_L^z) \end{pmatrix}. \end{aligned} \quad (5)$$

The angle parameters  $u_i^y$  and  $u_i^z$  can be given in terms of the angles in elevation  $\theta_i^t$  and angles in azimuth  $\phi_i^t$  as  $u_i^z = (2\pi d/\lambda) \cos \theta_i^t$ , and  $u_i^y = (2\pi d/\lambda) \sin \theta_i^t \sin \phi_i^t$ . The  $i$ -th column of  $\mathbf{A}_r$  is given by

$$\mathbf{a}_{r,i} = \mathbf{a}_{N_{ry}}(u_i^y) \odot \mathbf{a}_{N_{rz}}(u_i^z). \quad (6)$$

Similarly for an  $N_{ty} \times N_{tz}$  UPA at the transmitter with  $N_t = N_{ty}N_{tz}$ , the matrix  $\mathbf{A}_t$ , of size  $N_t \times L$ , is given by

$$\mathbf{A}_t = \mathbf{A}_{ty} \odot \mathbf{A}_{tz}, \quad (7)$$

where  $\mathbf{A}_{ty}$ , of size  $N_{ty} \times L$ , and  $\mathbf{A}_{tz}$ , of size  $N_{tz} \times L$ , are defined as

$$\begin{aligned} \mathbf{A}_{ty} &= \begin{pmatrix} \mathbf{a}_{N_{ty}}(p_1^y) & \mathbf{a}_{N_{ty}}(p_2^y) & \cdots & \mathbf{a}_{N_{ty}}(p_L^y) \end{pmatrix}, \\ \mathbf{A}_{tz} &= \begin{pmatrix} \mathbf{a}_{N_{tz}}(p_1^z) & \mathbf{a}_{N_{tz}}(p_2^z) & \cdots & \mathbf{a}_{N_{tz}}(p_L^z) \end{pmatrix}. \end{aligned} \quad (8)$$

The angle parameters  $p_i^z$  and  $p_i^y$  can be given in terms of the angles in elevation  $\theta_i^t$  and angles in azimuth  $\phi_i^t$  as  $p_i^z = (2\pi d/\lambda) \cos \theta_i^t$ , and  $p_i^y = (2\pi d/\lambda) \sin \theta_i^t \sin \phi_i^t$ . For the  $m$ -th training block, suppose the precoder and combiner are  $\mathbf{F}[m]$  and  $\mathbf{G}[m]$ , respectively. The numbers of RF chains at the transmitter and the receiver are, respectively,  $N_r^{RF}$  and  $N_t^{RF}$ , where  $N_r > N_r^{RF}$  and  $N_t > N_t^{RF}$ . The transmit precoding matrix  $\mathbf{F}[m] = (\mathbf{f}_{0,m} \mathbf{f}_{1,m} \cdots \mathbf{f}_{N_t^{RF}-1,m})$  is of dimensions  $N_t \times N_t^{RF}$  and the receive combining matrix  $\mathbf{G}[m] = (\mathbf{g}_{0,m} \mathbf{g}_{1,m} \cdots \mathbf{g}_{N_r^{RF}-1,m})$  is of dimensions  $N_r \times N_r^{RF}$ . Notice that we use only RF precoding and combining for angle estimation. Baseband precoder and combiner are not used and hence are not shown. Consider the transmission of training vectors in blocks of length  $N_t^{RF}$ ,  $\mathbf{S} = (\mathbf{s}_0 \mathbf{s}_1 \cdots \mathbf{s}_{N_t^{RF}-1})$ ,

<sup>1</sup>The elevation angle of a ray is the angle between the ray and the  $z$ -axis whereas the azimuth angle is the angle between the  $x$ -axis and the orthogonal projection of the ray on the  $xy$ -plane.

where  $\mathbf{s}_i$  is  $N_t^{RF} \times 1$ . The  $m$ -th received block is given by

$$\mathbf{Y}[m] = \mathbf{G}^H[m] \mathbf{H}[m] \mathbf{F}[m] \mathbf{S} + \mathbf{G}^H[m] \mathbf{N}[m], \quad (9)$$

where  $\mathbf{N}[m]$  is the  $N_r \times N_t^{RF}$  noise matrix. The noise is white Gaussian with zero mean and variance  $\sigma_n^2$ . We assume the channel stays the same during a block of  $N_t^{RF}$  vectors but the path gains can vary for different training blocks. For simplicity, we will use  $\mathbf{S} = \sigma_s \mathbf{I}_{N_t^{RF}}$ . In this case, we have

$$\mathbf{Y}[m] = \mathbf{G}^H[m] \mathbf{A}_r \mathbf{D}[m] \mathbf{A}_t^H \mathbf{F}[m] + \tilde{\mathbf{N}}[m] \quad (10)$$

for  $m = 0, 1, \dots, K-1$ , where  $K$  is the number of received blocks and  $\tilde{\mathbf{N}}[m] = \mathbf{G}^H[m] \mathbf{N}[m]$ .

In the following we review an angle estimation problem that is termed a harmonic retrieval problem in the literature. We also give a review of banded Toeplitz matrices that will be used later in Section III.

*A brief review of harmonic retrieval problem [37]:* Consider an antenna array of  $N_r$  elements. Suppose there are  $L$  sources. The  $m$ -th received vector is given by

$$\mathbf{y}[m] = \mathbf{A} \mathbf{d}[m] + \mathbf{n}[m], \quad \text{for } m = 0, 1, \dots, K-1,$$

where  $\mathbf{d}[m]$  is the  $L \times 1$  source vector,  $\mathbf{n}[m]$  is the channel noise,  $K$  is the number of received vectors, and  $\mathbf{A}$  is the antenna response matrix that is determined by AoAs and the antenna array. For a ULA array,  $\mathbf{A}$  is the same as  $\mathbf{A}_r$  in (2). The  $i$ -th column of  $\mathbf{A}$  is  $(1 e^{ju_i} \dots e^{j(N_r-1)u_i})^T$ , which has the so-called one-dimensional (1D) shift invariance property, i.e., the bottom  $(N_r - 1)$  elements identical to the top  $(N_r - 1)$  elements except for a scalar of  $e^{ju_i}$ . Define  $\mathbf{Y} = (\mathbf{y}[0] \mathbf{y}[1] \dots \mathbf{y}[K-1])$ . Then

$$\mathbf{Y} = \mathbf{A} \mathbf{S}_d + \mathbf{N}, \quad (11)$$

where  $\mathbf{S}_d = (\mathbf{d}[0] \mathbf{d}[1] \dots \mathbf{d}[K-1])$ , of size  $L \times K$ , consists of the source vectors. The estimation of the AoAs based on the received  $\mathbf{Y}$  is a harmonic retrieval problem that can be solved using many methods. For example, it can be solved using ESPRIT, thanks to the shift invariance property. If the ranks of  $\mathbf{A}$  and  $\mathbf{S}_d$  are both equal to  $L$ , then  $L$  paths can be identified. When a UPA array is used,  $\mathbf{A}$  is the Khatri-Rao product of two ULA antenna response matrices like that given in (4). The estimation of AoA becomes a 2D problem. More generally, when  $\mathbf{A}$  is the Khatri-Rao product of  $j$  ULA antenna response matrices, we have a  $j$ -dimensional harmonic retrieval problem. Each column of  $\mathbf{A}$  has the  $j$ -dimensional shift invariance property. Estimation algorithms for such problems have been extensively studied in the literature, e.g., [24], [25], [26], [27].

*A review of banded Toeplitz matrices:* In the presence of precoding and combining, the received vectors are no longer in the span of shift-invariant antenna response vectors and subspace based methods are not applicable. For ULA, 1D shift invariant property can be preserved if the precoder is lower triangular banded Toeplitz [22]. An  $N \times M$  lower triangular

banded Toeplitz matrix  $\mathbf{B}$  is given by

$$\mathbf{B} = \begin{pmatrix} b_0 & 0 & & 0 \\ b_1 & b_0 & & 0 \\ \vdots & \vdots & \ddots & \\ & & & b_0 \\ b_{N-M} & b_{N-M-1} & & \\ 0 & b_{N-M} & & \\ \vdots & & & \vdots \\ 0 & 0 & & b_{N-M} \end{pmatrix}, \quad (12)$$

Due to the Toeplitz structure, the matrix can be obtained from the vector  $\mathbf{b} = (b_0 \ b_1 \ \dots \ b_{N-M})^T$ , which will be referred to as the *prototype* of the matrix.

### III. ANGLE ESTIMATION USING BBT PRECODING AND COMBINING

Consider the received block of vectors in (10) with UPA arrays. The UPA receive antenna response matrix  $\mathbf{A}_r$  is the Khatri-Rao product  $\mathbf{A}_{r_y} \odot \mathbf{A}_{r_z}$  as given in (4) and  $\mathbf{A}_t = \mathbf{A}_{t_y} \odot \mathbf{A}_{t_z}$  as in (7). Suppose the number of RF chains at the receiver can be factorized as  $N_r^{RF} = N_{r_y}^{RF} N_{r_z}^{RF}$ . We propose to use a combiner of the following form

$$\mathbf{G}[m] = \mathbf{G}_y[m] \otimes \mathbf{G}_z[m], \quad (13)$$

where  $\mathbf{G}_y[m]$ , of size  $N_{r_y} \times N_{r_y}^{RF}$ , and  $\mathbf{G}_z[m]$ , of size  $N_{r_z} \times N_{r_z}^{RF}$ , are banded Toeplitz as in (12). As both  $\mathbf{G}_y[m]$  and  $\mathbf{G}_z[m]$  are banded Toeplitz, the combiner is banded block Toeplitz (BBT). Let the prototype of  $\mathbf{G}_y[m]$  be  $\mathbf{g}_{y,m}$ , then all the column vectors of  $\mathbf{G}_y[m]$  can be obtained from  $\mathbf{g}_{y,m}$ . Similarly all the column vectors of  $\mathbf{G}_z[m]$  can be obtained from its prototype  $\mathbf{g}_{z,m}$ . Define the discrete time Fourier transform of the coefficients of  $\mathbf{g}_{y,m}$  as

$$G_{y,m}(\omega) = \sum_{n=0}^{N_{r_y}-N_{r_y}^{RF}} [\mathbf{g}_{y,m}]_n e^{-j n \omega}$$

and the discrete time Fourier transform of  $\mathbf{g}_{z,m}$  as  $G_{z,m}(\omega) = \sum_{n=0}^{N_{r_z}-N_{r_z}^{RF}} [\mathbf{g}_{z,m}]_n e^{-j n \omega}$ . The received vectors in (10) are in the column space of  $\mathbf{G}^H[m] \mathbf{A}_r$ . When  $\mathbf{G}[m]$  is BBT, we can establish the following result.

*Lemma 1:* Suppose the combiner  $\mathbf{G}[m]$  is BBT as in (13). We have

$$\mathbf{G}^H[m] (\mathbf{A}_{r_y} \odot \mathbf{A}_{r_z}) = (\bar{\mathbf{A}}_{r_y} \odot \bar{\mathbf{A}}_{r_z}) \mathbf{D}_g^*[m], \quad (14)$$

where  $\mathbf{D}_g[m]$  is a diagonal matrix with

$$[\mathbf{D}_g[m]]_{i,i} = G_{y,m}(u_i^y) G_{z,m}(u_i^z).$$

$\bar{\mathbf{A}}_{r_y}$  is an  $N_{r_y}^{RF} \times L$  matrix whose  $i$ -th column is  $\mathbf{a}_{N_{r_y}^{RF}}(u_i^y)$  and  $\bar{\mathbf{A}}_{r_z}$  is  $N_{r_z}^{RF} \times L$  whose  $i$ -th column is  $\mathbf{a}_{N_{r_z}^{RF}}(u_i^z)$ .

*Proof:* See Appendix A. ■

Both  $\bar{\mathbf{A}}_{r_y}$  and  $\bar{\mathbf{A}}_{r_z}$  are ULA antenna response matrices. Lemma 1 means the received vectors are in the column space

of  $\mathbf{A}_{ry} \odot \mathbf{A}_{rz}$ , which consists of 2D shift-invariant antenna response vectors. For the transmitter, we also use a BBT matrix for precoding. Suppose the number of RF chains at the transmitter is the product  $N_t^{RF} = N_{ty}^{RF} N_{tz}^{RF}$ . The precoding matrix  $\mathbf{F}[m]$  is given by

$$\mathbf{F}[m] = \mathbf{F}_y[m] \otimes \mathbf{F}_z[m], \quad (15)$$

where the two matrices  $\mathbf{F}_y[m]$  and  $\mathbf{F}_z[m]$  are banded Toeplitz as in (12), of sizes  $N_{ty} \times N_{ty}^{RF}$ , and  $N_{tz} \times N_{tz}^{RF}$ , respectively. Let  $\mathbf{f}_{y,m}$  be the prototype of  $\mathbf{F}_y[m]$  and  $\mathbf{f}_{z,m}$  be the prototype of  $\mathbf{F}_z[m]$ . The Fourier transform of  $\mathbf{f}_{y,m}$  and  $\mathbf{f}_{z,m}$  are denoted as  $F_{y,m}(\omega)$  and  $F_{z,m}(\omega)$ , respectively. Similar to the result in Lemma 1, we have

$$(\mathbf{A}_{ry} \odot \mathbf{A}_{tz})^H \mathbf{F}[m] = \mathbf{D}_f[m] (\bar{\mathbf{A}}_{ry} \odot \bar{\mathbf{A}}_{tz})^H \quad (16)$$

where  $\mathbf{D}_f[m]$  is a diagonal matrix with  $[\mathbf{D}_f[m]]_{i,i} = F_{y,m}(p_i^y) F_{z,m}(p_i^z)$ . The matrix  $\bar{\mathbf{A}}_{ry}$  is  $N_{ty}^{RF} \times L$ , whose  $i$ -th column is  $\mathbf{a}_{N_{ty}^{RF}}(p_i^y)$ , and  $\bar{\mathbf{A}}_{tz}$  is  $N_{tz}^{RF} \times L$ , whose  $i$ -th column is  $\mathbf{a}_{N_{tz}^{RF}}(p_i^z)$ . Combining the results in (14) and (16), we have the following lemma.

*Lemma 2:* When both  $\mathbf{F}[m]$  and  $\mathbf{G}[m]$  are BBT, the  $m$ -th received block in (10) is given by

$$\mathbf{Y}[m] = \bar{\mathbf{A}}_r \mathbf{D}[m] \bar{\mathbf{A}}_t^H + \bar{\mathbf{N}}[m], \quad (17)$$

$$m = 0, 1, \dots, K - 1,$$

where  $\bar{\mathbf{A}}_r = \bar{\mathbf{A}}_{ry} \odot \bar{\mathbf{A}}_{rz}$ , and  $\bar{\mathbf{A}}_t = \bar{\mathbf{A}}_{ty} \odot \bar{\mathbf{A}}_{tz}$ . The matrix  $\mathbf{D}[m]$  is diagonal with the  $i$ -th diagonal element

$$[\mathbf{D}[m]]_{i,i} = \sigma_s \alpha_i[m] F_{y,m}(p_i^y) F_{z,m}(p_i^z) G_{y,m}^*(u_i^y) G_{z,m}^*(u_i^z). \quad (18)$$

Observe that  $\bar{\mathbf{A}}_r$  is the antenna response matrix of an  $N_{ry}^{RF} \times N_{rz}^{RF}$  UPA and  $\bar{\mathbf{A}}_t$  is the antenna response matrix of an  $N_{ty}^{RF} \times N_{tz}^{RF}$  UPA. The signal part  $\bar{\mathbf{A}}_r \mathbf{D}[m] \bar{\mathbf{A}}_t^H$  in (17) is the same as the received data of a system equipped with UPAs, but with no precoding and combining. The equivalent  $i$ -th path gain is  $[\mathbf{D}[m]]_{i,i}$ . The expression in (18) implies that we have partial control of the equivalent path gains through the precoder and combiner. We can introduce artificial fading to the equivalent path gains by using precoder and combiner that vary block by block. Consider the following prototype for  $\mathbf{G}_y[m]$ ,

$$[\mathbf{g}_{y,m}]_n = \gamma e^{j2\pi \frac{\varepsilon_{m,n}}{2^{N_Q}}}, \text{ for } n = 0, 1, \dots, N_{ry} - N_{ry}^{RF}, \quad (19)$$

where  $\varepsilon_{m,n}$  is randomly drawn from  $\{0, 1, \dots, 2^{N_Q} - 1\}$  with equal probability,  $N_Q$  is a chosen positive integer, and  $\gamma = \frac{1}{\sqrt{N_{ry} - N_{ry}^{RF} + 1}}$  is a normalizing factor so that the prototype has unit norm. We design the prototypes of  $\mathbf{G}_z[m]$ ,  $\mathbf{F}_y[m]$ , and  $\mathbf{F}_z[m]$  to have random coefficients similar to that in (19). Then the prototypes of different training blocks are independent. The equivalent path gains in (18) are time-varying even if the original path gains are time-invariant. Vectorizing the received matrix  $\mathbf{Y}[m]$  and calling it  $\mathbf{y}[m]$ , we have

$$\mathbf{y}[m] = \mathbf{A} \mathbf{d}[m] + \bar{\mathbf{n}}[m], \quad (20)$$

---

**Algorithm 1:** Proposed Angle Estimation.

---

- 1: Generate time-varying BBT precoders  $\mathbf{F}[0], \mathbf{F}[1], \dots, \mathbf{F}[K - 1]$  and combiners  $\mathbf{G}[0], \mathbf{G}[1], \dots, \mathbf{G}[K - 1]$  as in (13).
- 2: Apply precoder  $\mathbf{F}[m]$  and combiner  $\mathbf{G}[m]$  for the  $m$ th training block, where  $m = 0, 1, \dots, K - 1$ .
- 3: Collect  $K$  received block of vectors

$$\mathbf{Y}[m] = \bar{\mathbf{A}}_r \mathbf{D}[m] \bar{\mathbf{A}}_t^H + \bar{\mathbf{N}}[m],$$

and form the vectorized versions  $\mathbf{y}[m] = \text{vec}(\mathbf{Y}[m])$  for  $m = 0, 1, \dots, K - 1$ .

- 4: Construct the data matrix

$$\mathbf{Y}_{(nb)} = \begin{bmatrix} \mathbf{y}[0] & \mathbf{y}[1] & \dots & \mathbf{y}[K - 1] \end{bmatrix}.$$

- 5: Obtain estimates  $\hat{u}_1^y, \hat{u}_2^y, \dots, \hat{u}_L^y$ , and  $\hat{u}_1^z, \hat{u}_2^z, \dots, \hat{u}_L^z$  for AoAs and estimates  $\hat{p}_1^y, \hat{p}_2^y, \dots, \hat{p}_L^y$  and  $\hat{p}_1^z, \hat{p}_2^z, \dots, \hat{p}_L^z$  for AoDs from  $\mathbf{Y}_{(nb)}$  using harmonic-retrieval methods.
- 

where  $\mathbf{d}[m]$  is an  $L \times 1$  vector that consists of the diagonal elements of  $\mathbf{D}[m]$  and the equivalent antenna response matrix is

$$\mathbf{A} = \bar{\mathbf{A}}_t^* \odot \bar{\mathbf{A}}_r = \bar{\mathbf{A}}_{ty}^* \odot \bar{\mathbf{A}}_{tz}^* \odot \bar{\mathbf{A}}_{ry} \odot \bar{\mathbf{A}}_{rz}. \quad (21)$$

Collecting  $K$  training blocks, we have  $\mathbf{Y} = [\mathbf{y}[0] \ \mathbf{y}[1] \ \dots \ \mathbf{y}[K - 1]]$ , which can be written as

$$\mathbf{Y}_{(nb)} = \mathbf{A} \mathbf{S}_d + \mathbf{N}, \quad (22)$$

where  $\mathbf{S}_d = [\mathbf{d}[0] \ \mathbf{d}[1] \ \dots \ \mathbf{d}[K - 1]]$  and  $\mathbf{N} = [\bar{\mathbf{n}}[0] \ \bar{\mathbf{n}}[1] \ \dots \ \bar{\mathbf{n}}[K - 1]]$ . A subscript ‘(nb)’ has been added to indicate the narrowband case. We see that (22) is the same as the harmonic retrieval problem in (11) if we treat the time-varying  $\mathbf{d}[k]$  as  $L$  source signals. We say the problem in (22) is four-dimensional as the equivalent antenna response matrix  $\mathbf{A}$  in (21) is the Khatri-Rao product of four ULA antenna response matrices. Summarizing, we have the following theorem.

*Theorem 1:* Suppose the precoder  $\mathbf{F}[m]$  and combiner  $\mathbf{G}[m]$  are time-varying BBT with random prototypes as designed in (19). Then  $\mathbf{Y}_{(nb)}$  in (22) is equivalent to the received matrix of a time-varying channel, whether the channel is time-invariant or time-varying. The angles can be estimated using methods developed for harmonic retrieval problems.

An angle estimation algorithm based on Theorem 1 is given in Algorithm 1. If ULA are used at the transmitter and receivers,  $\bar{\mathbf{A}}_r$  reduces to  $\bar{\mathbf{A}}_{rz}$ ,  $\bar{\mathbf{A}}_t$  reduces to  $\bar{\mathbf{A}}_{tz}$ , and thus  $\mathbf{A}$  is the Khatri-Rao product of two ULA antenna response matrices. The problem in (22) becomes a 2D harmonic retrieval problem. When we use  $\mathbf{Y}_{(nb)}$  to estimate the channel, the number of identifiable paths is related to the size of  $\mathbf{Y}_{(nb)}$ , as shown in the following lemma.

*Lemma 3:* Consider the angle estimation problem in (22). We can identify  $L$  paths in the absence of noise if

$N_{ty}^{RF}$ ,  $N_{tz}^{RF}$ ,  $N_{ry}^{RF}$ , and  $N_{rz}^{RF}$  are at least 2 and

$$L \leq \min(N_{ty}^{RF} + N_{tz}^{RF} + N_{ry}^{RF} + N_{rz}^{RF} - 3, K). \quad (23)$$

*Proof:* See Appendix B.

The above results means that the number of training blocks  $K$  can be as small as  $L$ . When the transmitter and receiver are equipped with ULAs, we can use the proof in Appendix B to show that the condition in (23) becomes  $L \leq \min(N_t^{RF} + N_r^{RF} - 2, K)$ . Lemma 3 also means that  $K$  should be at least  $L$ . The overhead can be further reduced at the cost of a smaller measurement dimension by applying spatial smoothing [16]. After estimates of the angles are obtained, the path gains can be estimated using a least square approach. For the estimation of path gains, the receivers need to know the precoder used. For this, we can generate the random phase in (19) using a random number generator and the seed can be made available to the receiver ahead of time.

*Remarks:* In the above discussion, we have considered angle estimation for only point-to-point transmission. The proposed method can also be used for uplink multi-user hybrid mmWave MIMO systems. This can be done by designing training vectors as in [34]. The training vectors of different users are time-multiplexed so that at any instant, the received vector is from only one user. As a result, angle estimation for the multi-user case can be done for each user separately. Notice that the coefficients of the analog RF precoder are changed between two training blocks. To allow time for changing the values of the RF phase shifters, we can leave time between two training blocks and send out training blocks as fast as it is allowed. Time-varying precoders are also employed in [3], [4], and [12]. Also note that an BBT precoder contains some nonzero and zero coefficients. The nonzero coefficients can be implemented using  $N_Q$ -bit phase shifters. For the implementation of the zero coefficients, switches may be used [30].

#### IV. CHANNEL ESTIMATION FOR WIDEBAND TRANSMISSION

In the narrowband case the delays of different paths are small compared to the symbol period  $T$  and they are ignored. For wideband transmission, the delays need to be taken into consideration. In this section, we show that BBT precoding can be used for the joint estimation of angles and delays whether the channel is time-varying or time-invariant. Suppose the training time is divided into  $K$  training intervals. For the  $q$ -th training interval, a wideband geometric channel of  $L$  paths is of the form [18]

$$\mathbf{H}_q(t) = \mathbf{A}_r \mathbf{D}_q(t) \mathbf{A}_t^H, \quad (24)$$

where  $\mathbf{A}_r$  and  $\mathbf{A}_t$  are as in (4) and (7),  $\mathbf{D}_q(t)$  is an  $L \times L$  diagonal matrix with the  $i$ -th diagonal element given by

$$[\mathbf{D}_q(t)]_{i,i} = \alpha_i[q] p(t - \tau_i),$$

$\alpha_i[q]$  is the gain of the  $i$ th path gain in the  $q$ -th training interval,  $p(t)$  is the pulse shaping function, and  $\tau_i$  is the

delay associated with the  $i$ -th path. Suppose  $p(t)$  has an excess bandwidth with respect to the symbol rate  $\beta$  and a finite support  $[0, L_p T]$ . Let  $\tau_{max}$  be the maximum delay spread and  $N_c T = L_p T + \tau_{max}$ , then  $\mathbf{H}_q(t)$  is nonzero for  $t \in [0, N_c T]$ . The path gains remain the same in each training interval but can change to different values in different training intervals due to fading [29]. For the  $q$ -th training interval, the precoder and combiner are, respectively,  $\mathbf{F}[q]$  and  $\mathbf{G}[q]$ , which are used for the entire  $q$ -th training interval. Suppose  $M = N_c N_t^{RF}$  training vectors are transmitted in each training interval and the received signal is sampled with  $P$  times the symbol rate, i.e., the sampling period  $T_s = \frac{T}{P}$ . The received vector at the  $q$ -th training interval is given by

$$\begin{aligned} \mathbf{y}_q(t) = & \sum_{k=-(N_c-1)T_s}^n \mathbf{G}^H[q] \mathbf{H}_q(t - kT) \mathbf{F}[q] \mathbf{s}(k) \\ & + \mathbf{G}^H[q] \mathbf{n}_q(t), \end{aligned} \quad (25)$$

where  $t = 0, T_s, \dots, (MP - 1)T_s$ , and  $n = \lfloor \frac{t}{T_s} \rfloor$ . Consider the following training sequence

$$\mathbf{s}(k) = \begin{cases} \sigma_s \mathbf{e}_i, & k = iN_c, \\ 0, & k \neq iN_c, \end{cases} \quad \text{for } k = 0, 1, \dots, N_c N_t^{RF} - 1, \quad (26)$$

where  $[\mathbf{e}_i]_i = 1$  and  $[\mathbf{e}_i]_j = 0$  for  $j \neq i$ . That is,  $N_c - 1$  zero vectors are inserted between every two nonzero training vectors to prevent inter-symbol interference. The resulting training sequence for the  $q$ -th training interval is thus

$$\sigma_s \mathbf{e}_0 \mathbf{0} \cdots \mathbf{0} \sigma_s \mathbf{e}_1 \mathbf{0} \cdots \mathbf{0} \cdots \sigma_s \mathbf{e}_{N_c-1} \mathbf{0} \cdots \mathbf{0} \quad (27)$$

Let  $\bar{\mathbf{y}}_q(t) = [\mathbf{y}_q^T(t) \mathbf{y}_q^T(t + N_c T) \cdots \mathbf{y}_q^T(t + N_c(N_t^{RF} - 1)T)]^T$  for  $t = 0, T_s, \dots, (N_c P - 1)T_s$ . We collect all the received vectors in the  $q$ -th training interval into an  $N_r^{RF} N_t^{RF} \times N_c P$  matrix,

$$\tilde{\mathbf{Y}}[q] = \begin{pmatrix} \bar{\mathbf{y}}_q(0) & \bar{\mathbf{y}}_q(T_s) & \cdots & \bar{\mathbf{y}}_q((N_c P - 1)T_s) \end{pmatrix}.$$

It turns out that the column space of  $\tilde{\mathbf{Y}}[q]$  carries the angle information while the the row space of  $\tilde{\mathbf{Y}}[q]$  carries the delay information, as stated in the following lemma.

*Lemma 4:* Suppose the precoder  $\mathbf{F}[q]$  and combiner  $\mathbf{G}[q]$  are BBT as in (13) and the training sequence in (26) is used. Then  $\tilde{\mathbf{Y}}[q]$  can be expressed as

$$\tilde{\mathbf{Y}}[q] = \mathbf{A} \mathbf{D}[q] \mathbf{P}^T + \mathbf{N}[q], \quad (28)$$

where  $\mathbf{A}$  is as defined in (21),  $\mathbf{D}[q]$  is a diagonal matrix whose diagonal elements are as given in (18), and  $\mathbf{P} = (\mathbf{p}(\tau_1) \mathbf{p}(\tau_2) \cdots \mathbf{p}(\tau_L))$  is an  $N_c P \times L$  matrix with the  $i$ -th column given by  $\mathbf{p}(\tau_i) = [p(-\tau_i) p(T_s - \tau_i) \cdots p((N_c P - 1)T_s - \tau_i)]^T$ . The noise matrix  $\mathbf{N}[q]$  is given by  $\mathbf{N}[q] = (\mathbf{I}_{N_r^{RF}} \otimes \mathbf{G}^H[q]) [\tilde{\mathbf{n}}_q(0) \tilde{\mathbf{n}}_q(T_s) \cdots \tilde{\mathbf{n}}_q(N_c T - T_s)]$ , and

$$\tilde{\mathbf{n}}_q(t) = \begin{pmatrix} \mathbf{n}_q^T(t) & \mathbf{n}_q^T(t + N_c T) & \cdots & \mathbf{n}_q^T(t + N_c(N_t^{RF} - 1)T) \end{pmatrix}^T.$$

*Proof:* See Appendix C.  $\blacksquare$

The expression in (28) is free from inter-symbol interference. Using the training sequence in (26), we have transformed a wideband angle estimation problem to a narrowband problem. The information of angles is buried in the antenna response matrix  $\mathbf{A}$  while the information of delays is embedded in the matrix  $\mathbf{P}$ . We can further process  $\mathbf{P}$  to construct an antenna response matrix corresponding to the delays in a manner similar to [36]. To do this, let  $\mathbf{U}_o$  be the  $N'_c P \times N'_c P$  DFT matrix, where  $N'_c \geq N_c$ , and let  $\mathbf{U}$  be the  $N_c P \times N'_c P$  submatrix of  $\mathbf{U}_o$  that consists of the first  $N_c P$  rows. Observe that the  $i$ -th column of  $\mathbf{P}$  is  $\mathbf{p}(\tau_i)$ , which consists of samples of the pulse shaping function with sampling period  $T_s$ . The  $i$ -th row of  $\mathbf{P}^T \mathbf{U}$  contains the  $N'_c P$ -point DFT of  $\mathbf{p}^T(\tau_i)$ . We assume that the oversampling factor  $P \geq (1 + \beta)$  so that there is no aliasing when  $p(t)$  is sampled. In this case, each row of  $\mathbf{P}^T \mathbf{U}$  contains at most  $N'_c(1 + \beta)$  nonzero coefficients. Defined the  $N'_c P \times \bar{N}_d$  selection matrix

$$\mathbf{J}_d = \begin{bmatrix} \mathbf{0} & \mathbf{I}_{\lfloor \bar{N}_d/2 \rfloor} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{I}_{\lceil \bar{N}_d/2 \rceil} & \mathbf{0} \end{bmatrix}, \quad (29)$$

where  $\bar{N}_d \leq N'_c(1 + \beta)$ , then  $\mathbf{P}^T \mathbf{U} \mathbf{J}_d$  picks out only nonzero coefficients of  $\mathbf{P}^T \mathbf{U}$ . Let  $\tilde{\mathbf{p}} = \mathbf{U}^T \mathbf{p}(0)$ , where  $\mathbf{p}(0) = (p(0) p(T_s) \cdots p((N_c P - 1)T_s))^T$ . Consider the  $N_r^{RF} \times \bar{N}_d$  matrix  $\mathbf{Y}^u[q]$  that is defined as  $\mathbf{Y}^u[q] = \tilde{\mathbf{Y}}[q] \mathbf{U} \mathbf{J}_d$ . Then, it can be shown (Appendix D) that

$$\mathbf{Y}^u[q] = \mathbf{A} \mathbf{D}[q] \mathbf{D}_\xi \bar{\mathbf{A}}_d^T \mathbf{D}_p + \mathbf{N}^u[q], \quad (30)$$

where  $\mathbf{D}_\xi$  is an  $L \times L$  diagonal matrix with the  $i$ -th diagonal element give by

$$[\mathbf{D}_\xi]_{i,i} = e^{-i\xi_i \lceil \bar{N}_d/2 \rceil},$$

and  $\xi_i = -\frac{2\pi}{N'_c} \frac{\tau_i}{T}$ . The  $\bar{N}_d \times L$  matrix  $\bar{\mathbf{A}}_d$  is a ULA antenna response matrix with the  $i$ -th column vector  $\mathbf{a}_{\bar{N}_d}(\xi_i)$ ,  $\mathbf{D}_p$  is an  $\bar{N}_d \times \bar{N}_d$  invertible diagonal matrix whose diagonal entries are the elements of  $\tilde{\mathbf{p}}^T \mathbf{J}_d$ , and  $\mathbf{N}^u[q]$  is the corresponding noise matrix. The expression of  $\mathbf{Y}^u[q]$  in (30) contains  $\bar{\mathbf{A}}_d$ , which carries the delay information. We define

$$\mathbf{Y}'[q] = \mathbf{Y}^u[q] \mathbf{D}_p^{-1} = \mathbf{A} \mathbf{D}[q] \mathbf{D}_\xi \bar{\mathbf{A}}_d^T + \mathbf{N}'[q], \quad (31)$$

where  $\mathbf{N}'[q] = \mathbf{N}^u[q] \mathbf{D}_p^{-1}$ . Vectorizing  $\mathbf{Y}'[q]$ , we get  $\mathbf{y}'[q] = (\bar{\mathbf{A}}_d \odot \mathbf{A}) \mathbf{d}'[q] + \mathbf{n}'[q]$ , where  $\mathbf{n}'[q] = \text{vec}(\mathbf{N}'[q])$  and  $\mathbf{d}'[q]$ , consisting of the diagonal elements of  $\mathbf{D}[q] \mathbf{D}_\xi$ , is given by

$$[\mathbf{d}'[q]]_i = \sigma_s e^{-i\xi_i \lceil \bar{N}_d/2 \rceil} \alpha_i[q] F_{0,q}(p_i) G_{0,q}^*(u_i). \quad (32)$$

Then  $\mathbf{Y}_{(wb)} = [\mathbf{y}'[0] \mathbf{y}'[1] \cdots \mathbf{y}'[K-1]]$  can be expressed as

$$\mathbf{Y}_{(wb)} = (\bar{\mathbf{A}}_d \odot \mathbf{A}) \mathbf{S}_{d'} + \mathbf{N}', \quad (33)$$

where  $\mathbf{S}_{d'} = [\mathbf{d}'[0] \mathbf{d}'[1] \cdots \mathbf{d}'[K-1]]$ ,  $\mathbf{N}' = [\mathbf{n}'[0] \mathbf{n}'[1] \cdots \mathbf{n}'[K-1]]$  and we have added the subscript 'wb' to indicate the wideband case. The  $L \times 1$  vector  $\mathbf{d}'[q]$ , containing the equivalent path gains of the  $q$ -th training

interval, is partially controlled by the precoder and combiner. Artificial fading can be introduced by using BBT precoder and combiner that vary in different training intervals. Notice that (33) has the same form as the harmonic retrieval problem in (11). It is a five-dimensional harmonic retrieval problem as  $\bar{\mathbf{A}}_d \odot \mathbf{A}$  contains five ULA antenna response matrices. We can use, e.g., [27], to estimate the four types of angles and delays. When the transmitter and receiver are equipped with ULA instead of UPA,  $\bar{\mathbf{A}}_d \odot \mathbf{A}$  contains three ULA antenna response matrices and the harmonic retrieval problem is three-dimensional. Summarizing, we have the following result.

*Theorem 2:* Suppose we design the training sequence as in (26) and use BBT precoder and combiner with prototypes as in (13). Then,  $\mathbf{Y}_{(wb)}$  in (33) is equivalent to the received matrix of a time-varying channel with path gain vector  $\mathbf{d}'[q]$  for  $q = 0, 1, \dots, K-1$ . The angles and delays can be estimated as in a typical harmonic retrieval problem, whether the channel is time-invariant or time-varying.

Notice that (32) has a form similar to that in (18). However the index 'q' of  $\mathbf{d}'[q]$  in (32) is the index of training interval, whereas the index 'm' of  $\mathbf{d}[m]$  in (22), is the index of training block. The procedure for the proposed wideband estimation is given in Algorithm 2. Following a procedure similar to the proof of Lemma 2, we can show that  $L$  paths can be identified if

$$L \leq \min(N_{ry}^{RF} + N_{tz}^{RF} + N_{ry}^{RF} + N_{rz}^{RF} + \bar{N}_d - 4, K). \quad (34)$$

That is, the number of training intervals  $K$  can be as few as  $L$ . In the case ULAs are used, the condition in (34) becomes  $L \leq \min(N_t^{RF} + N_r^{RF} + \bar{N}_d - 2, K)$ .

*Remarks:* Oversampling has also been considered in [17] [16], [36] for joint estimation of angles and delays. However, these works require the path gains be time-varying. The time separation of the training intervals needs to be larger than the channel coherence time so that the channel is time-varying.

*Computational Complexity:* The complexity of the proposed angle estimation method depends on the algorithm used for harmonic retrieval. We analyze the complexity for the case when ESPRIT is used. We first form the received matrix  $\mathbf{Y}_{(nb)}$  as in (22) for the narrowband case, or  $\mathbf{Y}_{(wb)}$  (33) for the wideband case. Then we compute the empirical covariance matrix  $\frac{1}{K} \mathbf{Y}_{(nb)} \mathbf{Y}_{(nb)}^H$  (or  $\frac{1}{K} \mathbf{Y}_{(wb)} \mathbf{Y}_{(wb)}^H$ ), based on which the signal subspace is extracted and ESPRIT is applied. We list the complexity for narrowband and wideband cases in Table 1. For the step of subspace extraction, orthogonal iteration [37], [38] with  $i$  iterations is used in Table 1. We have included in Table 1 the complexity of angle pairing using the method in [33].

## V. SIMULATION RESULT

In the following examples, we use the mmWave channel model in (1) for narrowband transmission and (24) for wideband transmission, both with half-wavelength antenna

---

**Algorithm 2:** Joint Estimation of Angles and Delays for Wideband Channels.
 

---

- 1: Given the number of training intervals  $K$ , generate time-varying BBT precoders  $\mathbf{F}[0], \mathbf{F}[1], \dots, \mathbf{F}[K-1]$  and combiners  $\mathbf{G}[0], \mathbf{G}[1], \dots, \mathbf{G}[K-1]$  as in (13).
- 2: Apply precoder  $\mathbf{F}[m]$  and combiner  $\mathbf{G}[m]$  for the  $m$ th training interval, where  $m = 0, 1, \dots, K-1$ .
- 3: Stack the received training vectors and form  $\tilde{\mathbf{y}}_q(t) = [\mathbf{y}_q^T(t) \mathbf{y}_q^T(t + N_c T) \dots \mathbf{y}_q^T(t + N_c(N_t^{RF} - 1)T)]^T$ , where  $\mathbf{y}_q(t)$  is obtained by sampling the received vector signal as in (25).
- 4: Construct the data matrix of the  $q$ -th training interval as

$$\tilde{\mathbf{Y}}[q] = [\tilde{\mathbf{y}}_q(0) \tilde{\mathbf{y}}_q(T_s) \dots \tilde{\mathbf{y}}_q((N_c P - 1)T_s)].$$

- 5: Compute  $\mathbf{Y}^u[q] = \tilde{\mathbf{Y}}[q]\mathbf{U}\mathbf{J}_d$  and  $\mathbf{Y}'[q] = \mathbf{Y}^u[q]\mathbf{D}_p^{-1}$  as in (31).
- 6: Form the overall data matrix

$$\mathbf{Y}_{(wb)} = [\mathbf{y}'[0] \mathbf{y}'[1] \dots \mathbf{y}'[K-1]],$$

where  $\mathbf{y}'[q]$  is the vectorized version of  $\mathbf{Y}'[q]$ .

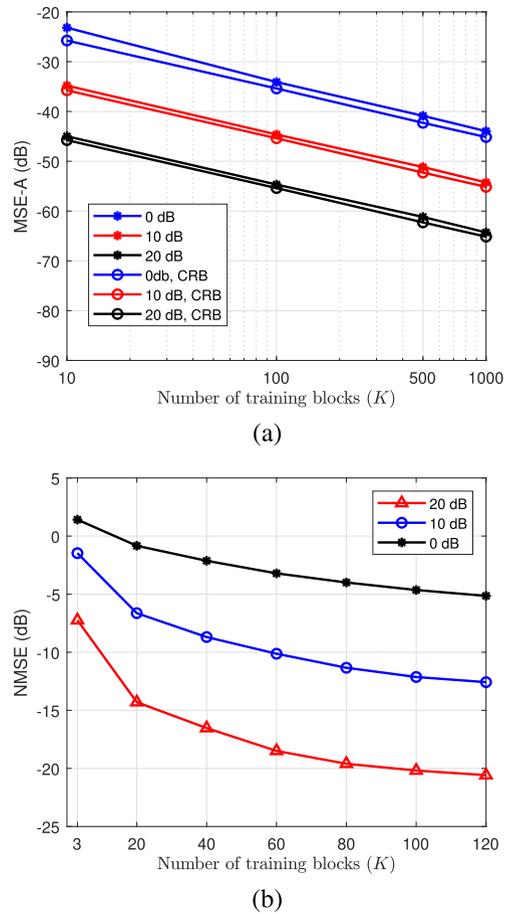
- 7: Obtain estimates of AoAs, estimates of AoDs and estimates of delays from  $\mathbf{Y}_{(wb)}$  using harmonic retrieval methods.
- 

spacing. For the wideband experiment in Example 5, practical mmWave channels with clustered paths are generated using the simulator in [39]. For time-varying channels, the variation of the path gains is modeled using the first-order Gauss-Markov process as in [20],

$$\alpha_i[m] = \epsilon \alpha_i[m-1] + \sqrt{1 - \epsilon^2} w_i[m-1]. \quad (35)$$

The channel correlation coefficient  $\epsilon$  is evaluated using Jake's model,  $\epsilon = J_0(2\pi f_d T_b)$ , where  $J_0(\cdot)$  is the zeroth order Bessel function of the first kind,  $f_d$  is the maximum Doppler frequency, and  $T_b$  is the time separation between two consecutive blocks. The random variable  $w_i[m-1]$ , independent of  $\alpha_i[m-1]$ , is a circular symmetric complex Gaussian random variable with unit variance. The prototypes of the random precoder and combiner are designed using  $N_Q = 4$  in (19).

*Example 1:* Consider the narrowband channel in (1) with ULA at the transmitter and receiver,  $N_t = N_r = 32$ , and  $N_t^{RF} = N_r^{RF} = 4$ . For a time-varying channel ( $\epsilon = 0$  in (35)) with  $L = 2$ , Fig. 2(a) shows the average mean squared error (MSE-A) of the angles  $\frac{1}{2L} \sum_{i=1}^L E[(\hat{p}_i - p_i)^2 + (\hat{u}_i - u_i)^2]$ . We have used  $\theta_1^t = 75^\circ$ ,  $\theta_2^t = 55^\circ$ ,  $\theta_1^r = 80^\circ$ , and  $\theta_2^r = 60^\circ$ . The plots are given for SNR  $\sigma_s^2/\sigma_n^2 = 0$  dB, 10 dB, and 20 dB. As a benchmark, the Cramér-Rao Bound (CRB) [40] of the estimation error has been plotted. The gap between the estimation error and the CRB bound is around 1 dB. Fig. 2(b) shows the normalized mean square error (NMSE) of the channel for a time-invariant channel ( $\epsilon = 1$  in (35)) with  $L = 3$  paths. The



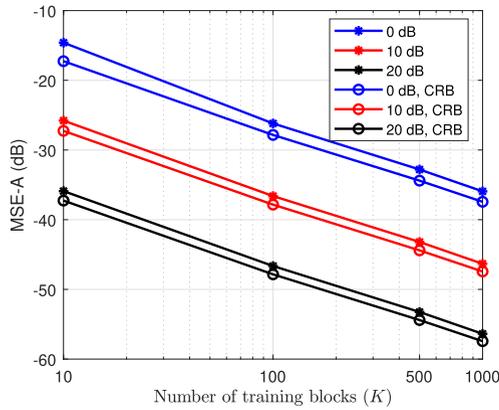
**FIGURE 2.** Estimation error for a narrowband channel with ULA for SNR  $\sigma_s^2/\sigma_n^2 = 0$  dB, 10 dB, and 20 dB. (a) MSE-A for a time-varying channel with  $L = 2$ ; (b) NMSE for a time-invariant channel with  $L = 3$ .

NMSE is defined as  $E[\|\hat{\mathbf{H}} - \mathbf{H}\|_F^2 / \|\mathbf{H}\|_F^2]$ , where  $\hat{\mathbf{H}}$  is the estimated channel matrix. The AoA and AoD are uniformly distributed over  $[-\pi, \pi]$  and SNR  $\sigma_s^2/\sigma_n^2 = 0$  dB, 10 dB, and 20 dB. We see that the NMSE decreases with  $K$  for a time-invariant channel. The proposed method works for both time-invariant and time-varying channels.

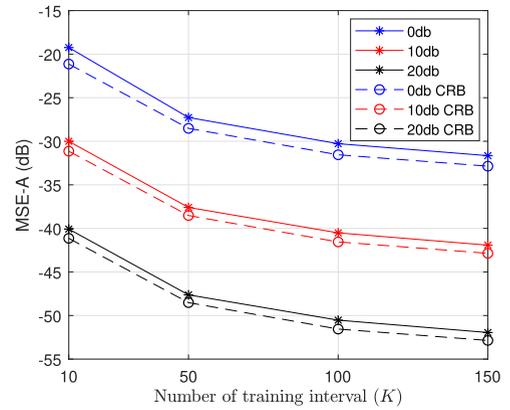
*Example 2:* Consider the case when UPAs are used at both the transmitter and receiver for the narrowband channel in (1) with  $N_t = N_r = 64$  ( $N_{ty} = N_{tz} = N_{ry} = N_{rz} = 8$ ) and  $N_t^{RF} = N_r^{RF} = 4$ . We use BBT precoder and combiner with  $N_{ty}^{RF} = N_{tz}^{RF} = N_{ry}^{RF} = N_{rz}^{RF} = 2$ . Fig. 3(a) shows the average mean squared error as a function of the number of training blocks. The channel is time-varying and there are  $L = 2$  paths with  $\theta_1^t = 75^\circ$ ,  $\theta_2^t = 55^\circ$ ,  $\phi_1^t = -20^\circ$ ,  $\phi_2^t = -40^\circ$ ,  $\theta_1^r = 80^\circ$ ,  $\theta_2^r = 60^\circ$ ,  $\phi_1^r = 0^\circ$ , and  $\phi_2^r = 20^\circ$ . The CRB [40] has also been plotted for comparison. In the above simulations we have assumed that knowledge of the number of paths  $L$  is available. When  $L$  is not known a priori, it can be estimated using model estimation [28]. Fig. 3(b) shows the NMSE of a time-invariant channel when  $L$  is known and when  $L$  is estimated using model estimation [28]. The angles are uniformly distributed

**TABLE 1** Complexity of the Proposed Algorithm

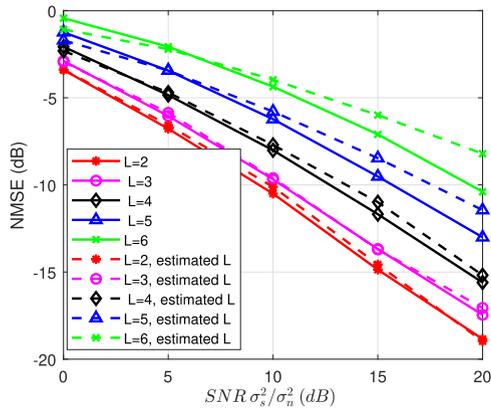
Dimension	Complexity (For signal subspace extraction, Orthogonal Iterations with $i$ iterations [37][38] is used.)
narrowband ULA	$O(N_t^{RF} N_r^{RF} ((2L^2 + L)i + 4L^2 + K + L^2) - 2L^2(N_t^{RF} + N_r^{RF}) + \frac{22}{3} L^3)$
wideband ULA	$O(N_t^{RF} N_r^{RF} \bar{N}_d ((2L^2 + L)i + 6L^2 + K + L^3) - 2L^2(N_t^{RF} N_r^{RF} + N_t^{RF} \bar{N}_d + N_r^{RF} \bar{N}_d) + 11L^3)$
narrowband UPA	$O(N_t^{RF} N_r^{RF} ((2L^2 + L)i + 8L^2 + K + L^4) - 2L^2(N_r^{RF} (N_{ty}^{RF} + N_{tz}^{RF}) + N_t^{RF} (N_{ry}^{RF} + N_{rz}^{RF})) + \frac{44}{3} L^2)$
wideband UPA	$O(N_t^{RF} N_r^{RF} \bar{N}_d ((2L^2 + L)i + 10L^2 + K + L^5) - 2L^2(N_r^{RF} \bar{N}_d (N_{ty}^{RF} + N_{tz}^{RF}) + N_t^{RF} \bar{N}_d (N_{ry}^{RF} + N_{rz}^{RF})) + N_t^{RF} N_r^{RF}) + \frac{55}{3} L^2)$



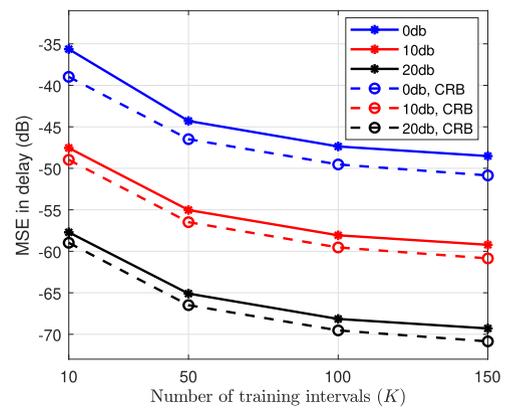
(a)



(a)



(b)



(b)

**FIGURE 3.** Performance for a narrowband channel with UPA. (a) MSE-A of a time-varying channel ( $L = 2$ ) for  $\text{SNR } \sigma_s^2 / \sigma_n^2 = 0 \text{ dB}, 10 \text{ dB},$  and  $20 \text{ dB}$ ; (b) NMSE for a time-invariant channel for  $L = 2, 3, \dots, 6$ .

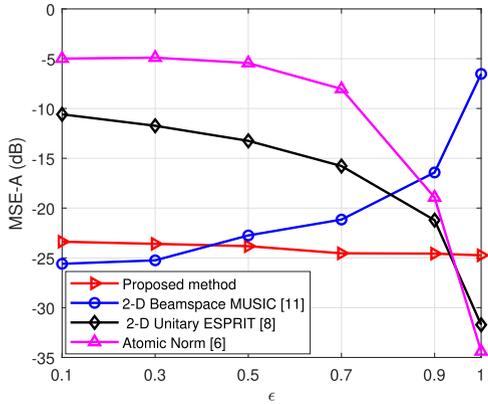
**FIGURE 4.** Estimation error for a wideband channel with UPA. MSE as a function of the number of training intervals for (a) angle and (b) delay.

over  $[-\pi, \pi]$  and  $K = 40$ . We see that having no prior knowledge of  $L$  leads to a small penalty in estimation error.

*Example 3:* Consider the wideband model in (24) when the transmitter and receiver are both equipped with UPA and the channel is time-varying with  $L = 2$ . The pulse shaping function  $p(t)$  is a raised cosine function with excess bandwidth  $\beta = 0.3$  and oversampling factor  $P = 2$ . The setup of antennas and RF chains is the same as that in Example 2, i.e., 64 antenna elements, and 4 RF chains. The AoAs and AoDs are the same as those in Fig. 3(a). The two path delays are  $\tau_1 = 1.3T$  and  $\tau_2 = 2.4T$ . Fig. 4(a) gives the MSE-A of the

angles and Fig. 4(b) the MSE of delays  $\frac{1}{L} \sum_{i=1}^L E[(\hat{\xi}_i - \xi_i)^2]$ , where  $\xi_i = -\frac{2\pi}{N_c} \frac{\tau_i}{T}$ , as given in (30). For delay estimation, we choose  $N'_c = 16$  in the construction of DFT matrix and  $\bar{N}_d = 16$  for the selection matrix in (29). It can be seen from Fig. 4 that both errors are close to the CRB [35], [40] for a wideband channel.

*Example 4:* In the above simulations, we have used the path gain model in (35) with the time correlation coefficient  $\epsilon = 0$ , or  $\epsilon = 1$ . In this example, we examine the estimation error as a function of  $\epsilon$  for a narrowband channel with  $L = 2$ ,  $N_t = N_r = 16$  and  $N_t^{RF} = N_r^{RF} = 4$ . ULAs are used

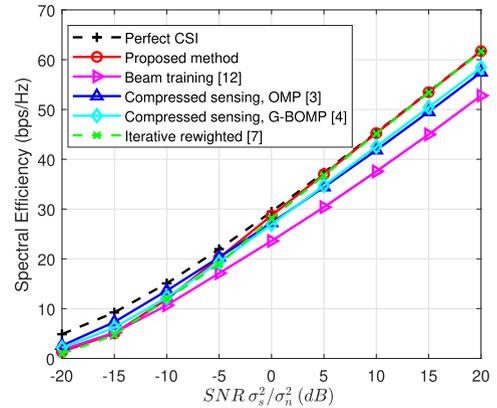


**FIGURE 5.** MSE-A as a function of channel correlation coefficient  $\epsilon$  for a narrowband channel with ULA when  $L = 2$  and  $\text{SNR } \sigma_s^2/\sigma_n^2 = 5$  dB.

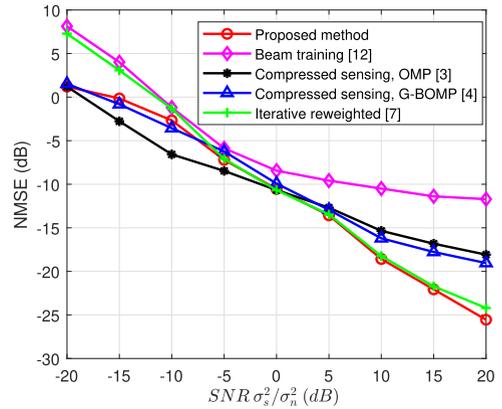
at both transmitter and receiver. Fig. 5 shows the MSE-A for  $p_1 = -0.2\pi$ ,  $p_2 = 0.1\pi$ ,  $u_1 = -0.1\pi$ ,  $u_2 = 0.2\pi$ ,  $\text{SNR } \sigma_s^2/\sigma_n^2 = 5$  dB, and  $K = 4$  training blocks. The result is compared with those of [6], [8], and [11]. The channel is assumed to be time-invariant during training in [6], [8], and assumed to be time-varying in [11]. We see that the error of [11] increases with  $\epsilon$  as the path gains become more static in time. Also, the performances of both [6] and [8] are degraded when  $\epsilon$  is small and the path gains vary faster. With the proposed method, the error is not significantly affected by  $\epsilon$ . Thanks to artificial fading introduced by time-varying BBT precoding, the proposed method is much more robust to channel correlation.

*Example 5:* In the data transmission to be followed after channel estimation, the precoder and combiner are designed according to the estimated channel, which results in degradation. In this example, the channel is time-invariant,  $N_t = N_r = 32$ ,  $N_t^{RF} = N_r^{RF} = 8$ ,  $L = 5$ , and ULAs are used at the transmitter and receiver. We consider narrowband transmission in Fig. 6 and wideband transmission Fig. 7.

For a narrowband scenario, Fig. 6(a) shows the achievable rate, which is computed using  $\log_2(\det(\mathbf{I}_{N_t^{RF}} + \rho \mathbf{R}^{-1} \mathbf{G}_{opt}^H \mathbf{H} \mathbf{F}_{opt} \mathbf{F}_{opt}^H \mathbf{H}^H \mathbf{G}_{opt}))$ , where  $\rho = \sigma_s^2/\sigma_n^2$ ,  $\mathbf{G}_{opt}$  consists of the dominant left singular vectors of  $\hat{\mathbf{H}}$ ,  $\mathbf{F}_{opt}$  consists of the dominant right singular vectors of  $\hat{\mathbf{H}}$ , and  $\mathbf{R} = \mathbf{G}_{opt}^H \mathbf{G}_{opt}$ . The result is compared with [3], [4], [7], [12]. In [12], hierarchical beam training is used for closed loop angle estimation; feedback is required before the training of the next stage. A total of 300 training vectors are used. For [3], [4], the angles are estimated using compressed sensing based methods with a dictionary size of 64 for both AoA and AoD. OMP algorithm is used in [3] and generalized block OMP (G-BOMP) algorithm is used in [4]. For the iterative reweighted method [7], the estimates are gridless. The number of training vectors for [3], [4] [7] and the proposed method is 128. Fig. 6(b) plots the NMSE for the five methods in Fig. 6(a). In terms of complexity, it is in the order of  $8.6 \times 10^4$  for the proposed method (computed from Table 1),  $4 \times 10^7$  for [12],  $4.2 \times 10^6$  for [3],  $2.7 \times 10^6$  for [4] (obtained from [14]), and



(a)



(b)

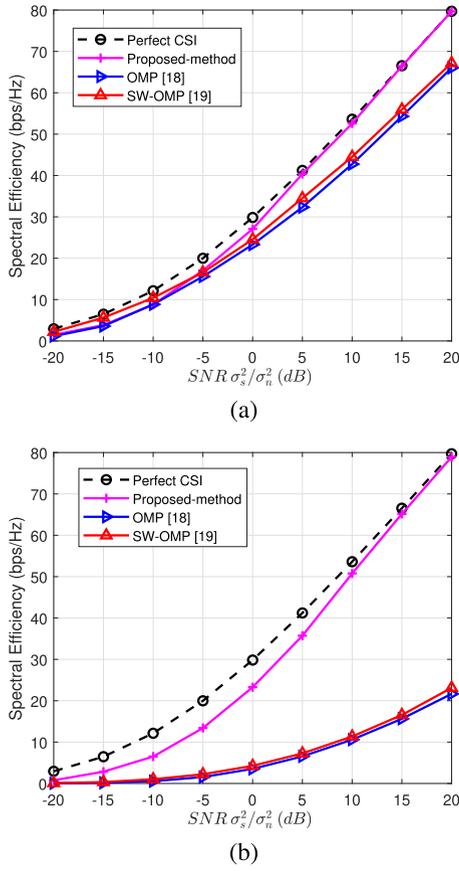
**FIGURE 6.** Performance comparison for a narrowband channel with  $L = 5$ ; (a) achievable rate and (b) NMSE.

$8.2 \times 10^7$  for [7]. The performance of the proposed method is comparable to that of the super resolution method in [7], but it enjoys a much lower complexity.

In Fig. 7, we consider joint estimation of delays and angles for a wideband channel. The mmWave channel simulator NYUSIM [39] is adopted to generate non line-of-sight channels with clustered paths. We have used carrier frequency 28 GHz, bandwidth 500 MHz and sampling frequency 1 GHz in the simulations. We plot in Fig. 7(a) the achievable rate, which is given by [19]

$$R = \frac{1}{N_f} \sum_{\kappa=0}^{N_f-1} \log_2(\det(\mathbf{I}_{N_t^{RF}} + \rho \mathbf{R}^{-1}(\kappa) \mathbf{G}_{opt}^H(\kappa) \mathbf{H}(\kappa) \mathbf{F}_{opt}(\kappa) \mathbf{F}_{opt}^H(\kappa) \mathbf{H}^H(\kappa) \mathbf{G}_{opt}(\kappa))),$$

where  $\mathbf{H}(\kappa) = \sum_{n=0}^{N_c-1} \mathbf{H}(nT) e^{-j \frac{2\pi\kappa}{N_f} n}$  is the  $\kappa$ th subcarrier channel, for  $\kappa = 0, 1, \dots, N_f - 1$ . We have used  $N_f = 128$ . The pulse shaping function  $p(t)$  is a raised cosine function with excess bandwidth  $\beta = 0.3$ . We choose  $P = 2$ ,  $N'_c = 64$  and  $\bar{N}_d = 64$  for delay estimation. Twelve training intervals, i.e., 5472 training vectors, are used in Fig. 7(a). Also shown



**FIGURE 7.** Achievable rate for a wideband mmWave channel with clustered paths, generated using NYUSIM [39]; (a) 12 training intervals and (b) one training interval.

In Fig. 7(a) are two compressed sensing based methods, [18] and [19]. For [18], [19], the dictionary size is 64 for both AoA and AoD, and the number of training vectors is 5580. In [18], [19], the actual delays are not estimated, but embedded in the subcarrier path gains. For the feedback of delay information, all the subcarrier path gains need to be fed back to the transmitter as explained in [14]. It can be seen from Fig. 7(a) that the proposed method can achieve a rate close to that of perfect CSI. In terms of complexity, it is in the order of  $2.1 \times 10^7$  for the proposed method (computed from Table 1), and  $1.5 \times 10^8$  for both [18] and [19]. In Fig. 7(b), we plot the achievable rate using the parameters in Fig. 7(a) except that the number of training intervals is reduced to one, i.e., 456 training vectors, for the proposed method. The number of training vectors used for [18], [19] is 558. The order of complexity for the proposed method is the same as in Fig. 7(a), i.e.,  $2.1 \times 10^7$ , and it is  $3.3 \times 10^7$  for both [18] and [19]. For the proposed method, using a small overhead causes some loss in the achievable rate, but the gap with the curve of perfect CSI narrows as SNR increases. In addition, our feedback overhead is much lower since the actual delays are estimated for each path. Therefore, our method enjoys a much lower complexity and feedback overhead by comparison.

## VI. CONCLUSION

In this paper, we propose an angle estimation method that can be applied to time-invariant or time-varying channels. We use BBT precoder and combiner with time-varying prototypes to convert the angle estimation problem for mmWave systems to a classic harmonic retrieval problem. Only a small training overhead is needed. We have considered narrowband transmission with ULA and UPA, and wideband transmission with ULA and UPA. For each case the angles can be estimated efficiently using methods developed in the literature for time-varying channels. We have demonstrated that gridless estimates can be obtained reliably with low complexity and low training overhead whether the channel is time-invariant or time-varying.

## APPENDIX A PROOF OF LEMMA 1

Using the property  $(\mathbf{B} \otimes \mathbf{C})(\mathbf{D} \odot \mathbf{E}) = (\mathbf{B}\mathbf{D}) \odot (\mathbf{C}\mathbf{E})$ , we have

$$\mathbf{G}^H[m](\mathbf{A}_{ry} \odot \mathbf{A}_{rz}) = (\mathbf{G}_y^H[m]\mathbf{A}_{ry}) \odot (\mathbf{G}_z^H[m]\mathbf{A}_{rz}). \quad (36)$$

The  $i$ -th column of  $\mathbf{G}_y^H[m]\mathbf{A}_{ry}$  is  $\mathbf{G}_y^H[m]\mathbf{a}_{N_{ry}}(u_i^y)$ . If  $\mathbf{G}_y[m]$  is banded Toeplitz as in (12), the  $k$ -th entry of  $\mathbf{G}_y^H[m]\mathbf{a}_{N_{ry}}(u_i^y)$  can be written as  $e^{jku_i^y} G_{y,m}^*(u_i^y)$ . Therefore  $\mathbf{G}_y^H[m]\mathbf{a}_{N_{ry}}(u_i^y)$  is  $G_{y,m}^*(u_i^y)\mathbf{a}_{N_{ry}}(u_i^y)$ . It follows that

$$\mathbf{G}_y^H[m]\mathbf{A}_{ry} = \bar{\mathbf{A}}_{ry}\mathbf{D}_{gy}^H[m],$$

where  $\mathbf{D}_{gy}[m]$  is an  $L \times L$  diagonal matrix with the  $i$ -th diagonal element given by  $[\mathbf{D}_{gy}[m]]_{i,i} = G_{y,m}(u_i^y)$ . Similarly, we have

$$\mathbf{G}_z^H[m]\mathbf{A}_{rz} = \bar{\mathbf{A}}_{rz}\mathbf{D}_{gz}^H[m],$$

where  $\mathbf{D}_{gz}[m]$  is an  $L \times L$  diagonal matrix with the  $i$ -th diagonal element given by  $[\mathbf{D}_{gz}[m]]_{i,i} = G_{z,m}(u_i^z)$ . Thus, we can express (36) as

$$\mathbf{G}^H[m](\mathbf{A}_{ry} \odot \mathbf{A}_{rz}) = (\bar{\mathbf{A}}_{ry}\mathbf{D}_{gy}^H[m]) \odot (\bar{\mathbf{A}}_{rz}\mathbf{D}_{gz}^H[m]).$$

As both  $\mathbf{D}_{gz}[m]$  and  $\mathbf{D}_{gy}[m]$  are diagonal matrices, we have  $\mathbf{G}^H[m](\mathbf{A}_{ry} \odot \mathbf{A}_{rz}) = (\bar{\mathbf{A}}_{ry} \odot \bar{\mathbf{A}}_{rz})(\mathbf{D}_{gz}^H[m]\mathbf{D}_{gy}^H[m])$ . Let  $\mathbf{D}_g[m] = \mathbf{D}_{gz}[m]\mathbf{D}_{gy}[m]$ , then we arrive at (14).

## APPENDIX B PROOF OF LEMMA 3

We first consider the detection of AoD parameters  $p_1^y, p_2^y, \dots, p_L^y$ . For the noiseless case, we have  $\mathbf{Y} = \mathbf{A}\mathbf{S}_d$ . Suppose ESPRIT is used. We compute the SVD to  $\mathbf{Y}$  to obtain an  $N_r^{RF} N_t^{RF} \times L$  matrix  $\mathbf{E}_s$  that has the same column space as  $\mathbf{A}$ . If  $N_{ty}^{RF} \geq 2$ , we can define the  $(N_{ty}^{RF} - 1) \times N_{ty}^{RF}$  selection matrices  $\mathbf{J}_1 = (\mathbf{I}_{N_{ty}^{RF}-1} \quad \mathbf{0})$  and  $\mathbf{J}_2 = (\mathbf{0} \quad \mathbf{I}_{N_{ty}^{RF}-1})$ . Also define

$$\mathbf{\Gamma}_m = \mathbf{J}_m \otimes \mathbf{I}_{N_{tz}^{RF}} \otimes \mathbf{I}_{N_{ry}^{RF}} \otimes \mathbf{I}_{N_{rz}^{RF}}, \quad m = 1, 2.$$

It is known that [27] if  $\mathbf{\Gamma}_1\mathbf{E}_s$  has full rank, then  $e^{jp_1^y}, e^{jp_2^y}, \dots, e^{jp_L^y}$  are the eigenvalues of  $(\mathbf{\Gamma}_1\mathbf{E}_s)^\dagger \mathbf{\Gamma}_2\mathbf{E}_s$ ,

where  $(\mathbf{\Gamma}_1 \mathbf{E}_s)^\dagger$  denotes a left inverse of  $\mathbf{\Gamma}_1 \mathbf{E}_s$ . Notice that  $\mathbf{\Gamma}_1 \mathbf{E}_s$  has full rank if the rank of  $\mathbf{\Gamma}_1 \mathbf{Y}$  is  $L$ , which is true if the ranks of  $\mathbf{S}_d$  and  $\mathbf{A}'$  are both  $L$ , where  $\mathbf{A}' = \bar{\mathbf{A}}'_{ty} \odot \bar{\mathbf{A}}'_{tz} \odot \bar{\mathbf{A}}'_{ry} \odot \bar{\mathbf{A}}'_{rz}$  and  $\bar{\mathbf{A}}'_{ty} = \mathbf{J}_1 \bar{\mathbf{A}}_{ty}$ . The rank of  $\mathbf{S}_d$  is equal to  $L$  if  $L \leq K$ . Observe that  $\mathbf{A}'$  is the Khatri-Rao product of four Vandermonde matrices. The rank of a Khatri-Rao product has been shown to be related to Kruskal rank (k-rank) [32]. The k-rank of a matrix  $\mathbf{X}$ , denoted by  $\text{krank}(\mathbf{X})$ , is defined as the maximum number  $k$  such that every  $k$  column vectors of  $\mathbf{X}$  are linearly independent, whereas the rank of  $\mathbf{X}$  is the maximal number of linearly independent columns. The k-rank of a matrix satisfies  $\text{rank}(\mathbf{X}) \geq \text{krank}(\mathbf{X})$  and for a Vandermonde matrix  $\mathbf{X}$ ,  $\text{krank}(\mathbf{X}) = \text{rank}(\mathbf{X})$ . It has been shown in [32] that, for two matrices  $\mathbf{N} \in \mathbb{C}^{n \times l}$  and  $\mathbf{M} \in \mathbb{C}^{m \times l}$ ,

$$\text{krank}(\mathbf{N} \odot \mathbf{M}) \geq \min(l, \text{krank}(\mathbf{N}) + \text{krank}(\mathbf{M}) - 1), \quad (37)$$

if  $\text{krank}(\mathbf{N}) \geq 1$  and  $\text{krank}(\mathbf{M}) \geq 1$ . This implies  $\text{krank}(\bar{\mathbf{A}}'_{ry} \odot \bar{\mathbf{A}}'_{rz}) \geq \min(L, \text{krank}(\bar{\mathbf{A}}'_{ry}) + \text{krank}(\bar{\mathbf{A}}'_{rz}) - 1)$ . Using this result in (37) repeatedly, we have

$$\text{rank}(\mathbf{A}') \geq \min(L, N_{ty}^{RF} + N_{tz}^{RF} + N_{ry}^{RF} + N_{rz}^{RF} - 3).$$

Therefore  $\text{rank}(\mathbf{A}') = L$  if  $L \leq N_{ty}^{RF} + N_{tz}^{RF} + N_{ry}^{RF} + N_{rz}^{RF} - 3$ . Combining this with the condition  $L \leq K$  yields (22). In a similar manner, we can verify that using the same condition, we can obtain AoD parameters  $p_1^z, p_2^z, \dots, p_L^z$  and also AoA parameters  $u_1^y, u_2^y, \dots, u_L^y$  and  $u_1^z, u_2^z, \dots, u_L^z$ . The result in Lemma 3 follows.

## APPENDIX C PROOF OF LEMMA 4

As the noise part is straightforward, we consider only the signal part, i.e., the noiseless case. When the training sequence is as in (26), the received vector is ISI-free and  $\mathbf{y}_q(t + iN_c T)$  can be written as  $\mathbf{y}_q(t + iN_c T) = \sigma_s \mathbf{G}^H[q] \mathbf{H}_q(t) \mathbf{F}[q] \mathbf{e}_i$ . Then

$$\bar{\mathbf{Y}}_q(t) = \left[ \mathbf{y}_q(t) \mathbf{y}_q(t + N_c T) \cdots \mathbf{y}_q(t + N_c(N_t^{RF} - 1)T) \right]$$

is equal to  $\sigma_s \mathbf{G}^H[q] \mathbf{H}_q(t) \mathbf{F}[q]$ . Using BBT precoder and combiner, we have  $\bar{\mathbf{Y}}_q(t) = \sigma_s \bar{\mathbf{A}}_r \mathbf{D}_g^H[q] \mathbf{D}_q(t) \mathbf{D}_f[q] \bar{\mathbf{A}}_t^H$ . Notice that the column vector that consists of the diagonal elements of  $\sigma_s \mathbf{D}_g^H[q] \mathbf{D}_q(t) \mathbf{D}_f[q]$  can be expressed as  $\mathbf{D}[q] \bar{\mathbf{p}}(t)$ , where  $\bar{\mathbf{p}}(t) = (p(t - \tau_1) p(t - \tau_2) \cdots p(t - \tau_L))^T$ . Thus, the vectorized version of  $\bar{\mathbf{Y}}_q(t)$  can be written as  $\bar{\mathbf{y}}_q(t) = \mathbf{A} \mathbf{D}[q] \bar{\mathbf{p}}(t)$ . Putting  $\bar{\mathbf{y}}_q(0), \bar{\mathbf{y}}_q(T_s), \dots, \bar{\mathbf{y}}_q((N_c P - 1)T_s)$  together, we obtain (28).

## APPENDIX D PROOF OF (30)

Let  $P(j\Omega)$  be the continuous time Fourier transform of  $p(t)$ , then the Fourier transform of  $p(t - \tau_i)$  is  $P(j\Omega)e^{-j\Omega\tau_i}$ . When the sampling rate is larger than the Nyquist rate and there is no aliasing, the discrete time Fourier transform of  $p(nT_s - \tau_i)$  is  $\sum_{n=-\infty}^{\infty} p(nT_s - \tau_i)e^{-jn\omega} = \frac{1}{T_s} P(j\frac{\omega}{T_s})e^{-j\frac{\omega}{T_s}\tau_i}$ , where  $|\omega| \leq \pi$ . By definition, the  $i$ -th row of  $\mathbf{P}^T \mathbf{U}$  is  $\mathbf{p}^T(\tau_i) \mathbf{U}$ . Notice that  $\mathbf{p}^T(\tau_i) \mathbf{U}$  computes the  $N_c P$ -point DFT of  $\mathbf{p}^T(\tau_i)$  as  $\mathbf{U}$  is a

submatrix of a DFT matrix. Thus

$$[\mathbf{p}^T(\tau_i) \mathbf{U}]_k = \frac{1}{T_s} P(j\frac{\omega}{T_s}) e^{-j\frac{\omega}{T_s}\tau_i} \Big|_{\omega = \frac{2\pi}{N_c P} k}.$$

It follows that

$$\mathbf{p}^T(\tau_i) \mathbf{U} \mathbf{J}_d = \left( e^{-j[\bar{N}_d/2]\xi_i} e^{-j([\bar{N}_d/2]-1)\xi_i} \cdots e^{j[\bar{N}_d/2]\xi_i} \right) \mathbf{D}_p.$$

Using the above expression, we arrive at (30).

## REFERENCES

- [1] S. Mumtaz, J. Rodriguez, and L. Dai, *Mmwave Massive MIMO: A Paradigm for 5G*, 1st ed. Norwell, MA, USA: Academic Press, 2016.
- [2] M. R. Akdeniz et al., "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.
- [3] J. Lee, G. Gil, and Y. H. Lee, "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2370–2386, Jun. 2016.
- [4] M. A. and A. P. Kannu, "Channel estimation strategies for multi-user mmWave systems," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5678–5690, Nov. 2018.
- [5] Y. Tsai, L. Zheng, and X. Wang, "Millimeter-wave beamformed full-dimensional MIMO channel estimation based on atomic norm minimization," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6150–6163, Dec. 2018.
- [6] H. Chu, L. Zheng, and X. Wang, "Super-resolution mmWave channel estimation for generalized spatial modulation systems," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 6, pp. 1336–1347, Oct. 2019.
- [7] C. Hu, L. Dai, T. Mir, Z. Gao, and J. Fang, "Super-resolution channel estimation for MmWave massive MIMO with hybrid precoding," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8954–8958, Sep. 2018.
- [8] A. Liao, Z. Gao, Y. Wu, H. Wang, and M.-S. Alouini, "2D unitary ESPRIT based super-resolution channel estimation for millimeter-wave massive MIMO with hybrid precoding," *IEEE Access*, vol. 5, pp. 24747–24757, 2017.
- [9] D. Fan et al., "Angle domain channel estimation in hybrid millimeter wave massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8165–8179, Dec. 2018.
- [10] Y. Wang, Y. Zhang, Z. Tian, G. Leus, and G. Zhang, "Super-resolution channel estimation for arbitrary arrays in hybrid millimeter-wave massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 5, pp. 947–960, Sep. 2019.
- [11] Z. Guo, X. Wang, and W. Heng, "Millimeter-wave channel estimation based on 2-D beamspace MUSIC method," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5384–5394, Aug. 2017.
- [12] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, pp. 831–846, Oct. 2014.
- [13] Y.-P. Lin and T.-M. Yang, "Random SBT precoding for angle estimation of mmWave massive MIMO systems using sparse arrays spacing," *IEEE Access*, vol. 8, pp. 163380–163393, 2020.
- [14] A. Liao, Z. Gao, H. Wang, S. Chen, M. Alouini, and H. Yin, "Closed-loop sparse channel estimation for wideband millimeter-wave full-dimensional MIMO systems," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8329–8345, Dec. 2019.
- [15] J. Zhang, D. Rakhimov, and M. Haardt, "Gridless channel estimation for hybrid mmWave MIMO systems via Tensor-ESPRIT algorithms in DFT beamspace," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 3, pp. 816–831, Apr. 2021.
- [16] J. Zhang and M. Haardt, "Channel estimation for hybrid multi-carrier mmWave MIMO systems using three-dimensional unitary esprit in DFT beamspace," *IEEE Int. Workshop Comput. Adv. Multi-Sensor Adaptive Process.*, 2017, pp. 1–5.
- [17] R. Shafin, L. Liu, Y. Li, A. Wang, and J. Zhang, "Angle and delay estimation for 3-D massive MIMO/FD-MIMO systems based on parametric channel modeling," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5370–5383, Aug. 2017.

- [18] K. Venugopal, A. Alkhateeb, N. González-Prelcic, and R. W. Heath, "Channel estimation for hybrid architecture-based wideband millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1996–2009, Sep. 2017.
- [19] J. Rodríguez-Fernández, N. González-Prelcic, K. Venugopal, and R. W. Heath, "Frequency-domain compressive channel estimation for frequency-selective hybrid millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2946–2960, May 2018.
- [20] S. Srivastava, P. Sharma, S. Dwivedi, A. K. Jagannatham, and L. Hanzo, "Fast block LMS based estimation of angularly sparse channels for single-carrier wideband millimeter wave hybrid MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 70, no. 1, pp. 666–681, Jan. 2021.
- [21] S. Srivastava, C. S. K. Patro, A. K. Jagannatham, and L. Hanzo, "Sparse Group-sparse, and online Bayesian learning aided channel estimation for doubly-selective mmWave hybrid MIMO OFDM systems," *IEEE Trans. Commun.*, vol. 69, no. 9, pp. 5843–5858, Sep. 2021.
- [22] P.-C. Chen and P. P. Vaidyanathan, "Convolutional beamspace for linear arrays," *IEEE Trans. Signal Process.*, vol. 68, pp. 5395–5410, 2020.
- [23] R. Roy and T. Kailath, "ESPRIT-Estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul. 1989.
- [24] L. Huang, Y. Wu, H. C. So, Y. Zhang, and L. Huang, "Multidimensional sinusoidal frequency estimation using subspace and projection separation approaches," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5536–5543, Oct. 2012.
- [25] J. Steinwandt, F. Roemer, M. Haardt, and G. D. Galdo, "Performance analysis of multi-dimensional ESPRIT-Type algorithms for arbitrary and strictly non-circular sources with spatial smoothing," *IEEE Trans. Signal Process.*, vol. 65, no. 9, pp. 2262–2276, May 2017.
- [26] E. Zhao, F. Zhang, D. Zhang, and S. Pan, "Three-dimensional multiple signal classification (3D-MUSIC) for super-resolution FMCW radar detection," in *Proc. IEEE MTT-S Int. Wireless Symp.*, 2019, pp. 1–3.
- [27] M. Haardt and J. A. Nossek, "Simultaneous Schur decomposition of several nonsymmetric matrices to achieve automatic pairing in multidimensional harmonic retrieval problems," *IEEE Trans. Signal Process.*, vol. 46, no. 1, pp. 161–169, Jan. 1998.
- [28] M. Wax and T. Kailath, "Determining the number of signals by information theoretic criteria," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, 1984, vol. 9, pp. 232–235.
- [29] S. Haghighatshoar and G. Caire, "Enhancing the estimation of mm-wave large array channels by exploiting spatio-temporal correlation and sparse scattering," *Proc. Int. ITG Workshop Smart Antennas*, 2016, pp. 1–7.
- [30] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R. W. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?," *IEEE Access*, vol. 4, pp. 247–267, 2016, doi: [10.1109/ACCESS.2015.2514261](https://doi.org/10.1109/ACCESS.2015.2514261).
- [31] Yuan-Pei Lin and Yu-Tai Chiew, "Beamforming aided angle estimation and feedback for hybrid mmWave systems," unpublished.
- [32] N. D. Sidiropoulos and R. Bro, "On the uniqueness of multilinear decomposition of n-way arrays," *J. Chemometrics*, vol. 14, no. 3, pp. 229–239, May 2000.
- [33] Y. Hua, "Estimating two-dimensional frequencies by matrix enhancement and matrix pencil," *IEEE Trans. Signal Process.*, vol. 40, no. 9, pp. 2267–2280, Sep. 1992.
- [34] D. Zhang, Y. Wang, Z. Su, and W. Xiang, "Millimeter wave channel estimation based on subspace fitting," *IEEE Access*, vol. 6, pp. 76126–76139, 2018, doi: [10.1109/ACCESS.2018.2884324](https://doi.org/10.1109/ACCESS.2018.2884324).
- [35] A.-J. van der Veen, M. C. Vanderveen, and A. Paulraj, "Joint angle and delay estimation using shift-invariance techniques," *IEEE Trans. Signal Process.*, vol. 46, no. 2, pp. 405–418, Feb. 1998.
- [36] M. C. Vanderveen, A.-J. Van der Veen, and A. Paulraj, "Estimation of multipath parameters in wireless communications," *IEEE Trans. Signal Process.*, vol. 46, no. 3, pp. 682–690, Mar. 1998.
- [37] M. Haardt, F. Roemer, and G. Del Galdo, "Higher-order SVD-Based subspace estimation to improve the parameter estimation accuracy in multidimensional harmonic retrieval problems," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3198–3213, Jul. 2008.
- [38] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd ed. Baltimore, MD, USA: The John Hopkins Univ. Press, 1989.
- [39] "NYUSIM Version 3.1 Now Available." NYU Wireless. [Online]. Available: <https://wireless.engineering.nyu.edu/nyusim/>
- [40] J. Steinwandt, F. Roemer, M. Haardt, and G. Del Galdo, "Deterministic Cramér-Rao bound for strictly non-circular sources and analytical analysis of the achievable gains," *IEEE Trans. Signal Process.*, vol. 64, no. 17, pp. 4417–4431, Sep. 2016.



research interests include signal processing for wireline and wireless communications.

**TING-MING YANG** (Senior Member, IEEE) was born in Taoyuan, Taiwan, in 1995. He received the B.S. degree in communication engineering from National Central University, Taoyuan, Taiwan, in 2018, and the M.S. degree from the Department of Electrical Engineering, National Yang Ming Chiao Tung University, Hsinchu, Taiwan, in 2020. Since 2021, he has been with MediaTek Inc., where he is currently an Engineer specializing in wireline SERDES transceiver architecture, DSP, and system engineering for multi-protocol SERDES. His



digital communications.

**YUAN-PEI LIN** (Senior Member, IEEE) was born in Taipei, Taiwan, 1970. She received the B.S. degree in control engineering from the National Chiao-Tung University, Hsinchu, Taiwan, in 1992, and the M.S. and Ph.D. degrees in electrical engineering from the California Institute of Technology, Pasadena, CA, USA, in 1993 and 1997, respectively. She joined the Department of Electrical and Control Engineering of National Chiao-Tung University in 1997. Her research interests include wireless communications and signal processing for