Bayesian Bilinear Inference for Joint Channel Tracking and Data Detection in Millimeter-Wave MIMO Systems

Takumi Takahashi, *Member, IEEE*, Hiroki Iimori, *Member, IEEE*, Koji Ishibashi, *Senior Member, IEEE*, Shinsuke Ibi, *Member, IEEE*, and Giuseppe Thadeu Freitas de Abreu, *Senior Member, IEEE*

***Abstract*— We propose a novel joint channel tracking and data detection (JCTDD) scheme to combat the channel aging phenomenon typical of millimeter-wave (mmWave) multiple-input multiple-output (MIMO) communication systems in high-mobility scenarios. The contribution aims to significantly reduce the communication overhead required to estimate time-varying mmWave channels by leveraging a Bayesian message passing framework based on Gaussian approximation, to jointly perform channel tracking (CT) and data detection (DD). The proposed method can be interpreted as an extension of the Kalman filter-based two-stage tracking mechanism to a Bayesian bilinear inference (BBI)-based joint channel and data estimation (JCDE) framework, featuring the ability to predict future channel state information (CSI) from both reference and payload signals by using an auto-regressive (AR) model describing the time variability of mmWave channel as a state transition model in a bilinear inference algorithm. The resulting JCTDD scheme allows us to track the symbol-by-symbol time variation of channels without embedding additional pilots, leaving any added redundancy to be exploited for channel coding, dramatically improving system performance. The efficacy of the proposed method is confirmed by computer simulations, which show that the proposed method not only significantly outperforms the state-of-the-art (SotA) but also approaches the performance of an idealized Genie-aided scheme.**

***Index Terms*— mmWave communications, data detection, channel aging, channel tracking, Bayesian bilinear inference.**

I. Introduction

MILLIMETER-WAVE (mmWave) technology is a promising solution to provide high data rate and low latency wireless communications services by exploiting the available wide spectrum in the 30 GHz to 300 GHz range [1], [2], [3], and has therefore been adopted in various recently-developed systems including 5G cellular and Wi-Fi 6. In order to fully realize the potential of mmWave communications, however, propagation losses must be effectively mitigated via beamforming techniques that require accurate channel state information (CSI).

Various effective beamforming schemes have therefore been proposed that leverage the statistical properties of mmWave wireless channels, such as their low-rankness and sparsity [4], [5], [6], [7]. These works assume, however, that the estimated mmWave channel remains constant over the duration of data transmission, while in practice the high directivity of mmWave signals makes their propagation highly susceptible to changes in the environment, such as mobility, reflections from obstacles, and the weather, which can result in significant variations in the channel characteristics [8], [9], [10], [11], [12], [13]. In particular, static or quasi-static fading assumptions are not valid for high-mobility scenarios [14], [15], [16] in time-varying mmWave wireless channels.

A primary challenge for wireless communications over time-varying mmWave multiple-input multiple-output (MIMO) channels is the increase in the communication overhead required for CSI acquisition [17], [18], [19], [20], [21], [22], [23]. In sharp beamforming, which uses a large number of antenna elements to concentrate power over a limited angular spread, it has been shown that even millisecond-scale fluctuations in the surrounding environment can cause the beam to lose its integrity with respect to the channel, causing significant performance degradation [13]. However, since reliable channel estimation (CE) inherently involves a pilot/training overhead that increases proportional to the number of antennas, frequent CE to maintain accurate CSI knowledge may result in non-negligible losses in spectrum efficiency.

To tackle this channel aging issue, several channel estimation/tracking (CE/CT) methods have been proposed for time-varying mmWave wireless communications. In [20], for instance, a two-stage channel tracking algorithm for downlink mmWave multi-user MIMO systems over time-varying channels was proposed under the assumption that the time variations of angles of arrival/departure (AoA/AoD) are much slower than the dynamics of the channel gains. For such

Manuscript received 27 June 2023; revised 11 November 2023 and 3 February 2024; accepted 13 March 2024. Date of publication 26 March 2024; date of current version 12 September 2024. This work was supported in part by Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant JP23K13335, Grant JP22H01483, and Grant JP21H01332. The associate editor coordinating the review of this article and approving it for publication was W. Chen. *(Corresponding author: Takumi Takahashi.)*

Takumi Takahashi is with the Graduate School of Engineering, Osaka University, Suita 565-0871, Japan (e-mail: takahashi@comm.eng.osaka-u.ac.jp).

Hiroki Iimori is with Ericsson Research, Yokohama, Kanagawa 220-0012, Japan (e-mail: h.iimori@ieee.org).

Koji Ishibashi is with the Advanced Wireless and Communication Research Center (AWCC), The University of Electro-Communications, Chofu-shi, Tokyo 182-8585, Japan (e-mail: koji@ieee.org).

Shinsuke Ibi is with the Faculty of Science and Engineering, Doshisha University, Kyotanabe 610-0394, Japan (e-mail: sibi@mail.doshisha.ac.jp).

Giuseppe Thadeu Freitas de Abreu is with the School of Computer Science and Engineering, Constructor University, 28759 Bremen, Germany (e-mail: gabreu@constructor.university).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2024.3379122.

Digital Object Identifier 10.1109/TWC.2024.3379122

a scenario, a novel pilot transmission frame was designed, where angle parameters are estimated at the beginning of the transmission frame while channel gains are tracked using shorter pilot sequences periodically inserted multiple times within the transmission frame. By separating the phenomena according to their time-varying rates and estimating each corresponding parameter separately, the required pilot overhead can be reduced. A similar separation and two-stage transmission frame were also employed in [21] and [22], in which the time-varying CE problem is formulated as a tensor decomposition problem solved by the least square (LS) approach, while taking Doppler shifts into consideration.

Another solution is the expectation maximization-based channel tracking (CT) method [23], which exploits the low-rank property of the mmWave MIMO channel. By making the assumption that the channel is quasi-static for a given time, and varies from block-by-block, this method achieves time-varying CE by tracking the model parameters of the sparse virtual channel with the Kalman filter (KF). The more realistic symbol-by-symbol time-varying nature of wireless channels, *i.e.*, channel aging, is often modeled as an auto-regressive (AR) process [24], [25], [26], [27], [28], but even here, the KF-based CT approach can still be utilized [25], [26], although machine learning (ML)-aided counterparts are also emerging, examples of which are the schemes in [26] and [27], where a deep neural network (DNN) was employed to perform channel prediction.

From the theoretical aspect, a rigorous analysis of the signal-to-interference-plus-noise ratio (SINR) and mean square error (MSE) performance of a system under time-variant Rayleigh fading was presented in [28]. After deriving a quasi-closed form of the average SINR leveraging random matrix theory, the importance of taking the auto-correlation coefficients into consideration in algorithm design was clarified.

We highlight that all the aforementioned studies considered CE/CT based only on reference signaling, *e.g.*, pilot sequences, which suggests that further overhead reduction can be achieved if estimated data symbols can be exploited as soft references for channel estimation. The latter approach is hereafter referred to as joint channel tracking and data detection (JCTDD).

A similar idea has been employed in the design of conventional joint channel and data estimation (JCDE) algorithms based on the Bayesian bilinear inference (BBI) framework [29], [30], [31], [32], [33] and has been shown, both in theory and implementation, to achieve high performance in a variety of wireless communication systems. Indeed, the authors have also recently shown that the BBI framework can be employed in the design of grant-free (GF) access schemes [34], joint activity and channel estimation (JACE) for extra-large MIMO (XL-MIMO) systems [35], and JCDE of cell-free massive MIMO (CF-mMIMO) systems [36], all of which were demonstrated to outperform earlier state-of-the-art (SotA) methods. Although the tracking aspect of the problem dramatically increases its difficulty, because both channel coefficients and data symbols are time-varying at the scale of symbol duration, the BBI framework at the base of SotA JCDE algorithms is also a key ingredient of the JCTDD algorithm introduced here as a novel solution to overcome the challenges related to time-varying mmWave MIMO systems. The idea of an extension of the BBI framework to the JCTDD problem requires elaborate algorithmic additions/modifications and is therefore not straightforward, which must incorporate a channel state transition model while appropriately combining messages to account for channel aging.

With that in mind, we extend the BBI framework to the JCTDD problem, theoretically enabling the compensation of channel aging effect without any additional pilots, via the alternation of the proposed CT and data detection (DD) solutions. This indicates that the processing of the proposed algorithm is completed inside the DD block and does not require any changes in frame format, higher-level functions, or protocols. The objective of this article is to provide new insight into the efficacy of the BBI framework in the mitigation of the time-varying nature of mmWave wireless communication channels. Specifically, we contribute with a customized BBI-based JCTDD receiver design to address the aforementioned inherent problems of time-varying mmWave MIMO systems, including CT and DD in the uplink.

Our contribution, summarized below, is fundamentally algorithmic, such that our results are presented in terms of the normalized mean square errors (NMSEs) of estimated quantities and bit error rates (BERs) computed from actual constellation points:

- A novel JCTDD algorithm is presented, wherein a BBI-based channel prediction (CP)-aided message passing is performed while exploiting the AR model describing the time-varying nature of mmWave wireless channels as the state transition model. The basic concept of the method is inherited from the two-stage tracking mechanism of the KF, and consists of a CP phase that generates prior estimates of current and future CSI based on past channel estimates, and a JCDE phase that generates posterior estimates of channel coefficients and data symbols based on those prior estimates, which are repeated sequentially. The main challenge to achieve JCTDD is, however, that only one snapshot is available for a given channel realization, since both the channel coefficients and the data symbols are time-varying, and the diversity gain across the time dimension cannot be obtained by applying the JCDE algorithm as-is. This is solved by also incorporating a CP mechanism based on the state transition model into the process to generate *beliefs* that propagate along edges on the factor graph (FG) in order to extract information about the current CSI from past and future observations. In other words, unlike the KF, here the prediction mechanism is present in both phases. An appropriate sequential update rule is also designed to ensure that the proposed JCTDD algorithm successfully captures the time variability of the channel.
- A new insight into the design of transceivers for time-varying mmWave MIMO systems is given, in that it is shown that, thanks to the BBI-based JCTDD algorithm, the addition of pilots within a transmission frame is not a suboptimal approach to exploit redundancy in the combat of time-varying channels, since the latter

is outperformed by channel coding. To elaborate, our receiver design is found to capitalize on available degrees of freedom (DoF) via its BBI framework, which enables highly accurate CT and DD without any additional pilots, even in relatively high-mobility communication scenarios with relative speed of 45 [kmph] or higher. In fact, this algorithmic feature allows the proposed method to operate based on the Bayesian framework in different transmission frame formats (*e.g.*, frame configurations with additional pilots), utilizing the additional available knowledge as prior information, without any change in the algorithm structure. Simulation results show that the proposed method approaches the ideal genie-aided performance using only an initial CSI estimated at the beginning of the frame, and channel coding thereafter.

In addition to the aforementioned contributions, we highlight that, to the best of our knowledge, there is no previous work in the literature offering a systematic JCTDD receiver design for time-varying mmWave MIMO communication systems via Bayesian bilinear inference.

Notation: The following notation is used throughout, unless when otherwise specified. Sets of counting, integer, real, and complex numbers are denoted by $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$, and $\mathbb{C}$, respectively. Vectors and matrices are denoted in lower- and upper-case bold-face fonts, respectively. The conjugate, transpose, and conjugate transpose operators are denoted by $(\cdot)^*$, $(\cdot)^{\mathsf{T}}$, and $(\cdot)^{\mathsf{H}}$, respectively. The expectation of a given random variable is denoted by $\mathbb{E}[\cdot]$. Random variables and their outcomes are denoted in sans serif and italic fonts, respectively, as in a and a, such that the conditional probability density function (PDF) and the conditional expectation of the outcome a of a, given the occurrence b of b are respectively denoted by $p_{\mathsf{a}|\mathsf{b}}(a|b)$ and $\mathbb{E}_{\mathsf{a}|\mathsf{b}}[a|b]$. The real and imaginary parts of a complex quantity are respectively denoted by $\Re\{\cdot\}$ and $\Im\{\cdot\}$. The $a \times a$ square identity matrix is denoted by $\boldsymbol{I}_a$. The (i,j)-th element, the j-th column vector, and the sub-matrix consisting of rows a_1 to a_2 and columns b_1 to b_2 of a matrix $\boldsymbol{A}$ are respectively denoted by $[\boldsymbol{A}]_{i,j}$, $[\boldsymbol{A}]_{:,j}$ and $[\boldsymbol{A}]_{a_1:a_2,b_1:b_2}$. The complex Gaussian distribution with mean a and variance b is denoted by $\mathcal{CN}(a,b)$. The notation $a \sim p_{\mathsf{a}}(a)$ indicates that a variable a obeys $p_{\mathsf{a}}(a)$. The symbols $\propto$ and $\otimes$ are used to denote proportionality and Kronecker product, respectively. Finally, following the style of related literatures [32], [33], [34], [35], and [36], we omit the integrating variable differential from all integrals and use the notation $\sum_{i\neq j}^{I} a_i \triangleq \sum_{i=1}^{I} a_i - a_j$, for brevity.

II. System Model

A. Channel Model

Consider a mmWave MIMO system with N_{TX} transmit (TX) antennas and N_{RX} receive (RX) antennas that simultaneously supports elevation and azimuth beamforming. Following existing literatures [37], [38], [39], and [40], the well-known clustered mmWave channel model with L clusters, each having C_ℓ rays with $\ell \in \mathcal{L} \triangleq \{1,2,\ldots,L\}$ is considered, so as to capture the sparsely scattered nature of mmWave channels. The corresponding $N_{\mathrm{RX}} \times N_{\mathrm{TX}}$ channel matrix can then be expressed as [41]:

$$\acute{\boldsymbol{H}} = \frac{1}{\sqrt{L}} \sum_{\ell=1}^{L} \sum_{c=1}^{C_\ell} \frac{\sigma_{\ell,c}}{\sqrt{C_\ell}} \underbrace{\boldsymbol{a}_{N_{\mathrm{RX}}}\left(\theta_{\ell,c}^{\mathrm{RX}}, \phi_{\ell,c}^{\mathrm{RX}}\right) \boldsymbol{a}_{N_{\mathrm{TX}}}^{\mathsf{H}}\left(\theta_{\ell,c}^{\mathrm{TX}}, \phi_{\ell,c}^{\mathrm{TX}}\right)}_{\triangleq \acute{\boldsymbol{A}}_{\ell,c}\text{: Array response matrix}}, \tag{1}$$

where $\theta_{\ell,c}^{\mathrm{RX}}$ and $\theta_{\ell,c}^{\mathrm{TX}}$ are respectively the elevation angle of arrival (AoA) and angle of departure (AoD) corresponding to the c-th ray of the ℓ-th scattering cluster, $\phi_{\ell,c}^{\mathrm{RX}}$ and $\phi_{\ell,c}^{\mathrm{TX}}$ are respectively the azimuth AoA and AoD corresponding to the c-th ray of the ℓ-th scattering cluster, and $\sigma_{\ell,c} \sim \mathcal{CN}(0,1)$ is a small-scale fading coefficient. Assuming that both the transmitter and receiver are equipped with uniformly spaced planar antenna arrays (UPAs), each with half-wavelength antenna element separations along the elevation-and-azimuth-axis, the array response vector in (1) is then given by

$$\boldsymbol{a}_P(\theta,\phi) \triangleq \boldsymbol{c}_{\sqrt{P}}(\sin(\theta)\cos(\phi)) \otimes \boldsymbol{c}_{\sqrt{P}}(\cos(\theta)), \tag{2}$$

with

$$\boldsymbol{c}_{\sqrt{P}}(\upsilon) \triangleq \frac{1}{\sqrt{P}} \left[1, e^{\mathrm{j}\pi\upsilon}, \ldots, e^{\mathrm{j}\pi(\sqrt{P}-1)\upsilon}\right]^{\mathsf{T}}. \tag{3}$$

The block fading assumption that the channel remains constant during the transmission of several data symbols holds for low-mobility telecommunication systems, but for mmWave orthogonal frequency-division multiplexing (OFDM)-MIMO systems, especially in high-mobility scenarios, the channel aging phenomenon cannot be ignored and this assumption often does not hold [14], [15], [16].

To represent the OFDM symbol-by-symbol time-varying nature of wireless channels, *i.e.*, channel aging, in high-mobility mmWave communication systems based on the channel model described above, at the discrete time index k, the time-varying characteristics of small-scale fading coefficients $\sigma_{\ell,c}$ in (1) can be expressed by using AR model as

$$\sigma_{\ell,c}[k] = \begin{cases} \mathcal{CN}(0,1) & k = 0, \\ r\sigma_{\ell,c}[k-1] + \sqrt{1-r^2}\omega_{\ell,c}[k] & k \in \mathbb{N}, \end{cases} \tag{4}$$

where $\omega_{\ell,c}[k] \sim \mathcal{CN}(0,1)$ is the time-varying component and r is the correlation parameter, *i.e.*, the AR coefficient of frequency-domain channels experienced by two consecutive OFDM symbols (corresponding to two consecutive discrete time indices), where the time correlation function is defined as

$$\rho(\Delta k) \triangleq \frac{\mathbb{E}\left[\sigma_{\ell,c}^*[k]\sigma_{\ell,c}[k+\Delta k]\right]}{\mathbb{E}\left[|\sigma_{\ell,c}[k]|^2\right]} = r^{\Delta k}, \tag{5}$$

with a discrete time-lag $\Delta k \in \mathbb{Z}$.

Consider a representative value of r in an OFDM system. According to [42], [43], and [44], the coherence time interval within which the channel auto-correlation function is above 0.5 is given by

$$T_{\mathrm{c}} \triangleq 0.423 \cdot \frac{v_{\mathrm{c}}}{v} \cdot \frac{1}{f_{\mathrm{c}}}, \tag{6}$$

where v (≥ 0) [m/s] denotes the relative velocity between transmitter and receiver, $v_{\rm c}$ [m/s] denotes the speed of light, and $f_{\rm c}$ [Hz] denotes the carrier frequency, respectively.

Given that the OFDM system operates with a discrete Fourier transform (DFT) of size $N_{\rm DFT}$ and a sampling rate $f_{\rm s}$ [Hz], the duration of each OFDM symbol is given by [16]

$$T_{\rm s} \triangleq N_{\rm DFT} \cdot \frac{(1+N_{\rm GI})}{f_{\rm s}} = (1+N_{\rm GI})\frac{1}{\Delta f}, \tag{7}$$

where $N_{\rm GI}$ is the guard interval and $\Delta f \triangleq f_{\rm s}/N_{\rm DFT}$ is the subcarrier spacing [45], [46].[1]

From (5)–(7), the maximum number of OFDM symbols within the coherence time is given by $K_{\max} \triangleq \lfloor T_{\rm c}/T_{\rm s} \rfloor$, and the time correlation parameter r used in (4) can then be calculated as[2]

$$r = \exp\left[\frac{\ln(0.5)}{K_{\max}}\right]. \tag{8}$$

From (1) and (4), denoting the number of OFDM symbols during data transmission by $K \leq K_{\max}$, such that the discrete time index $k \in \mathcal{K} \triangleq \{0,1,\ldots,K-1\}$, the aged frequency-domain channel at the k-th discrete time instance can be expressed as [15] and [16]

$$\begin{aligned}\acute{\boldsymbol{H}}[k] &= \frac{1}{\sqrt{L}}\sum_{l=1}^{L}\sum_{c=1}^{C_\ell}\frac{\sigma_{\ell,c}[k]}{\sqrt{C_\ell}}\acute{\boldsymbol{A}}_{\ell,c} \\ &= r^{k'}\acute{\boldsymbol{H}}[k-k'] \\ &\quad + \sqrt{\frac{1-r^2}{L}}\sum_{l=1}^{L}\sum_{c=1}^{C_\ell}\frac{\acute{\boldsymbol{A}}_{\ell,c}}{\sqrt{C_\ell}}\underbrace{\sum_{\kappa=0}^{k'-1} r^{\kappa}\omega_{\ell,c}[k-\kappa]}_{\triangleq \beta_{\ell,c}(k,k')} \\ &= r^{k}\acute{\boldsymbol{H}}[0] + \underbrace{\sqrt{\frac{1-r^2}{L}}\sum_{l=1}^{L}\sum_{c=1}^{C_\ell}\frac{\beta_{\ell,c}(k,k)}{\sqrt{C_\ell}}\acute{\boldsymbol{A}}_{\ell,c}}_{\text{Errors due to channel aging}}, \end{aligned} \tag{9}$$

where $\acute{\boldsymbol{H}}[0]$ is the initial CSI (estimated from the most recent pilots).

Similarly as in [37], it is assumed that the AoAs and AoDs are constant during the OFDM frame interval, while the channel coefficient may change rapidly. From the Gaussian-PDF multiplication rule [29], the cumulative change in small-scale fading due to channel aging, *i.e.*, $\beta_{\ell,c}(k,k')$, follows a Gaussian distribution with

$$\mathbb{E}\left[\beta_{\ell,c}(k,k')\right] = 0, \tag{10a}$$

$$\begin{aligned}\mathbb{E}\left[\left|\beta_{\ell,c}(k,k')\right|^2\right] &= \mathbb{E}\left[\left|\sum_{\kappa=0}^{k'-1} r^{\kappa}\omega_{\ell,c}[k-\kappa]\right|^2\right] \\ &= \sum_{\kappa=0}^{k'-1} r^{2\kappa}\underbrace{\mathbb{E}\left[\left|\omega_{\ell,c}[k-\kappa]\right|^2\right]}_{=1} \\ &= \frac{1-r^{2k'}}{1-r^2} \triangleq \phi^{\beta}_{\ell,c}(k'). \end{aligned} \tag{10b}$$

B. Signal Model

Based on the time-varying mmWave channel model described in the previous subsection, we describe the RX signal model of a point-to-point OFDM-MIMO communication system with a fully-digital TX/RX beamforming architecture. Denoting the TX and RX beamforming matrices respectively by $\boldsymbol{F}_{\rm TX} \in \mathbb{C}^{N_{\rm TX}\times M}$ ($M \leq N_{\rm TX}$) and $\boldsymbol{F}_{\rm RX} \in \mathbb{C}^{N_{\rm RX}\times N}$ ($N \leq N_{\rm RX}$), after cyclic prefix removal and DFT processing at the receiver, the frequency-domain RX vector of the f-th subcarrier at the k-th discrete time instance can be expressed as the following linear regression model

$$\begin{aligned}&\boldsymbol{y}_f[k] \\ &\triangleq [y_{f,1}[k],\ldots,y_{f,n}[k],\ldots,y_{f,N}[k]]^{\mathsf{T}} \\ &= \boldsymbol{F}_{\rm RX}^{\mathsf{H}}\acute{\boldsymbol{H}}_f[k]\boldsymbol{F}_{\rm TX}\boldsymbol{x}_f[k] + \boldsymbol{z}_f[k] = \boldsymbol{H}_f[k]\boldsymbol{x}_f[k] + \boldsymbol{z}_f[k], \end{aligned} \tag{11}$$

where $\boldsymbol{x}_f[k] \triangleq [x_{f,1}[k],\ldots,x_{f,m}[k],\ldots,x_{f,M}[k]]^{\mathsf{T}}$ is a TX vector, $\boldsymbol{H}[k] \triangleq [\boldsymbol{h}_{f,1}[k],\ldots,\boldsymbol{h}_{f,m}[k],\ldots,\boldsymbol{h}_{f,M}[k]] \triangleq \boldsymbol{F}_{\rm RX}^{\mathsf{H}}\acute{\boldsymbol{H}}_f[k]\boldsymbol{F}_{\rm TX} \in \mathbb{C}^{N\times M}$ is an effective channel matrix in the beam domain, and $\boldsymbol{z}_f[k] \triangleq [z_{f,1}[k],\ldots,z_{f,n}[k],\ldots,z_{f,N}[k]]^{\mathsf{T}} \sim \mathcal{CN}(0, N_0\boldsymbol{I}_N)$ is an independent and identically distributed (i.i.d.) circularly symmetric additive white Gaussian noise (AWGN) vector. Note that large-scale fading factors such as path-loss between the transmitter and receiver are estimated in advance and their effects are eliminated in (11). Hereafter, we will focus our discussion on a single subcarrier and therefore omit the subcarrier index f for notational simplicity.

In similarity to (9) in Subsection II-A, the effective beam-domain channel can be expressed as

$$\begin{aligned}\boldsymbol{H}[k] &= r^{k'}\boldsymbol{H}[k-k'] + \sqrt{\frac{1-r^2}{L}}\sum_{l=1}^{L}\sum_{c=1}^{C_\ell}\frac{\beta_{\ell,c}(k,k')}{\sqrt{C_\ell}}\boldsymbol{A}_{\ell,c} \\ &= r^{k}\boldsymbol{H}[0] + \sqrt{\frac{1-r^2}{L}}\sum_{l=1}^{L}\sum_{c=1}^{C_\ell}\frac{\beta_{\ell,c}(k,k)}{\sqrt{C_\ell}}\boldsymbol{A}_{\ell,c}, \end{aligned} \tag{12}$$

where $\boldsymbol{A}_{\ell,c} \triangleq \boldsymbol{F}_{\rm RX}^{\mathsf{H}}\acute{\boldsymbol{A}}_{\ell,c}\boldsymbol{F}_{\rm TX} \in \mathbb{C}^{N\times M}$ and $\boldsymbol{h}_m[k] \triangleq [h_{1m}[k],\ldots,h_{nm}[k],\ldots,h_{Nm}[k]]^{\mathsf{T}}$. For the sake of readability and future convenience, the second-order statistics of errors due to channel aging are computed based on (12) and summarized in Appendix.

C. Static TX/RX Beamforming Strategy

In practice, mmWave communications using TX/RX beamforming are realized by a two-step process: beam training to form static beams based on predesigned codebook and/or long-term channel statistics, and CT to maintain acceptable communication quality by tracking instantaneous channel variations due to small-scale fading [18], [19], [20], [21], [22].

[1] It is assumed that the time variability of channels during $T_{\rm s}$ is negligibly small and that there is no effect of inter-subcarrier interference.

[2] The term inside the exponential in equation (8) is always non-positive, since $\ln(0.5) < 0$, indicating that the correlation parameter r becomes smaller as $T_{\rm c}$ decreases (v and $f_{\rm c}$ increase) and $T_{\rm s}$ increases (Δf decreases), as intuited. When oversampling is performed, r becomes smaller as expected because $T_{\rm s}$ is inversely proportional to the sampling rate $f_{\rm s}$.

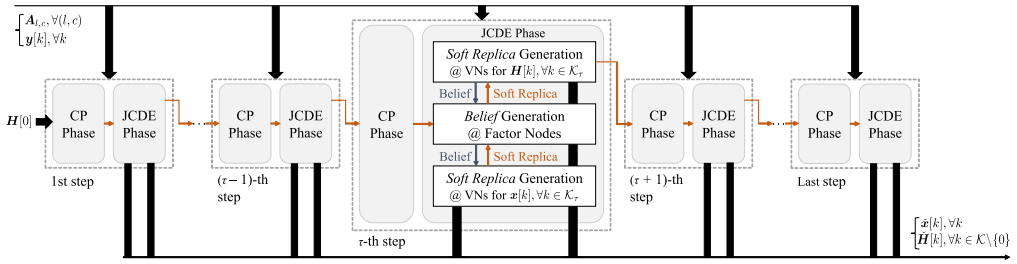


Fig. 1. Schematic of the belief propagation process employed in the proposed bilinear inference algorithm.

In this article, under the assumption that the beam training is conducted ideally, the optimal static digital beamforming is performed based on the singular value decomposition (SVD) of the initial CSI $\acute{\boldsymbol{H}}[0] = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathsf{H}}$ as[3]

$$\boldsymbol{F}_{\mathrm{TX}} = [\boldsymbol{V}]_{:,1:M}\,, \quad \boldsymbol{F}_{\mathrm{RX}} = [\boldsymbol{U}]_{:,1:N}\,. \tag{13}$$

If the block fading assumption holds, interference-free SVD-based MIMO communication can be achieved. However, in the presence of channel aging, the equivalent channel matrix after equalization, *i.e.*, $\boldsymbol{H}[k]^{\mathsf{H}}\boldsymbol{H}[k]$, has off-diagonal components for $k \in \mathcal{K}\setminus\{0\}$ due to mismatch between the beam and channel, resulting in non-negligible performance degradation as k increases.

The goal of the article can now be stated concisely: design a novel JCTDD algorithm that enables the receiver to simultaneously detect the intended TX vectors $\boldsymbol{x}[k], \forall k$, and accurately estimate/track the effective channel matrices $\boldsymbol{H}[k], \forall k \in \mathcal{K}\setminus\{0\}$, out of the RX vectors $\boldsymbol{y}[k], \forall k$, given in (11), the initial effective channel matrix $\boldsymbol{H}[0] = \boldsymbol{F}_{\mathrm{RX}}^{\mathsf{H}}\acute{\boldsymbol{H}}[0]\boldsymbol{F}_{\mathrm{TX}}$ estimated as a result of beam training, and the long-term channel statistics that change only slowly, *i.e.*, $\boldsymbol{A}_{\ell,c}, \forall(\ell,c)$. Unless otherwise specified, the TX/RX beamformers given by (13) shall be used.

III. Bayesian Joint Channel Tracking and Data Detection

In this section, a CP-aided message passing algorithm based on the BBI framework to design JCTDD receivers for time-varying mmWave MIMO architectures is described. To capture the time variability of wireless channels due to channel aging, the proposed algorithm has two phases: the CP phase and the JCDE phase. By iterating between these phases alternately, as illustrated in Fig. 1, the unknown data vectors $\boldsymbol{x}[k], \forall k$, and effective channel matrices $\boldsymbol{H}[k], \forall k \in \mathcal{K}\setminus\{0\}$, can be sequentially estimated, given the knowledge of RX vectors $\boldsymbol{y}[k], \forall k$, the known (estimated) effective channel matrix $\boldsymbol{H}[0]$, and the effective array response matrices $\boldsymbol{A}_{\ell,c}, \forall(\ell,c)$, obtained from long-term observation at the receiver.

For later convenience, the basic process consisting of one CP phase and one subsequent JCDE phase is defined as one *step*, and the set of discrete-time indices corresponding to the variables to be estimated in the process of the τ-th step, denoted by $\mathcal{K}_\tau \subset \mathcal{K}$, is designed flexibly according to the intended sequential processing. For example, in this problem setting, where the knowledge of $\boldsymbol{H}[0]$ is given, it is reasonable to perform estimation from $k = 0$ to $k = K-1$ in that order. In addition, since it is assumed that the relative velocity between the transmitter and receiver is constant during the OFDM frame interval, $\mathcal{K}_\tau$ is generated under a static rule with respect to τ. Following these remarks, the *active* set for each step of the algorithm is given by

$$\mathcal{K}_\tau \triangleq \left\{ k \in \mathcal{K} \,\middle|\, (\tau - D)\,W \le k \le \tau W - 1 \right\}, \quad \forall \tau \in \{1, 2, \ldots, \tau_{\max}\}\,, \tag{14}$$

with $\tau_{\max} = \frac{K}{W} + D - 1$, where $W \in \mathbb{N}$ is the number of new indices added to the active set for each step, while $D \in \mathbb{N}$ is a "window-sliding" parameter that determines the number of variables estimated simultaneously for each step. In view of (14), it will prove convenient to define the set $\mathcal{K}_\tau^+$, with $|\mathcal{K}_\tau^+| = \{0, W\}$, of discrete-time indices to be added to the active set between each step. For an intuitive understanding, Table I lists the sets $\mathcal{K}_\tau$ and $\mathcal{K}_\tau^+$ at each step when we set $(K, W, D) = (64, 2, 3)$ as an example.

Let us next shift our focus to each phase at the τ-th step. In the CP phase, the channels estimated in the previous $(\tau-1)$-th step are used to generate an initial estimate of channels corresponding to the indices newly added to the active set in the current τ-th step, *i.e.*, $\boldsymbol{H}[k], \forall k \in \mathcal{K}_\tau^+$. It is worth noting that these predicted channels are *prior* estimates that do not include information from the observations at $k \in \mathcal{K}_\tau^+$. In turn, in the JCDE phase, estimates of $\boldsymbol{x}[k], \forall k \in \mathcal{K}_\tau$, and $\boldsymbol{H}[k], \forall k \in \mathcal{K}_\tau$, are computed employing the bilinear Gaussian belief propagation (BiGaBP) approach, using the knowledge of $\boldsymbol{y}[k], \forall k \in \mathcal{K}_\tau$, of the predicted channels obtained in the $(\tau-1)$-th CP phase, and of the estimates of $\boldsymbol{x}[k], k \in \mathcal{K}_{\tau-1}$, obtained in the $(\tau-1)$-th JCDE phase. However, only the DD mechanism is based on the probabilistic data association (PDA) approach [48], which suppresses the interference using inverse matrices while taking into account the spatial correlation among fading coefficients.

The bilinear inference in the JCDE phase is performed by exchanging *beliefs* (*i.e.*, likelihood information reflecting detection reliability) and *soft replicas* (*i.e.*, tentative estimates) on the tripartite FG consisting of factor nodes (FNs) and two variable nodes (VNs), which correspond to the channel

[3]In case of hybrid beamforming architectures [47], where analog beamformers perform port reduction based on beam search methods without channel estimation, the SVD is calculated based on the effective channel matrix comprising of the actual channel and the analog beamformer. The proposed method described later is still compatible with such scenarios, since it can track the effective beam-domain channel matrix for a fixed analog beamformer. Notice that the SVD beamfomer on $\acute{\boldsymbol{H}}[0]$ as used in this article serves as the upper bound of such analog beam search-based alternatives.

TABLE I

EXAMPLES OF $\mathcal{K}_\tau$ AND $\mathcal{K}_\tau^+$ FOR THE CASE OF $(K, W, D) = (64, 2, 3)$

τ	$\mathcal{K}_\tau$	$\mathcal{K}_\tau^+$	τ	$\mathcal{K}_\tau$	$\mathcal{K}_\tau^+$
1	$\{0,1\}$	$\{0,1\}$	$\vdots$	$\vdots$	$\vdots$
2	$\{0,1,2,3\}$	$\{2,3\}$	30	$\{54,55,56,57,58,59\}$	$\{58,59\}$
3	$\{0,1,2,3,4,5\}$	$\{4,5\}$	31	$\{56,57,58,59,60,61\}$	$\{60,61\}$
4	$\{2,3,4,5,6,7\}$	$\{6,7\}$	32	$\{58,59,60,61,62,63\}$	$\{62,63\}$
5	$\{4,5,6,7,8,9\}$	$\{8,9\}$	33	$\{60,61,62,63\}$	$\emptyset$
$\vdots$	$\vdots$	$\vdots$	34	$\{62,63\}$	$\emptyset$

coefficients and TX data symbols, respectively. We remark that although the algorithm of the JCDE phase is designed based on the existing BiGaBP [33] and PDA approaches [48], both the belief generation in the FNs, which incorporate a new prediction mechanism to capture the time variation of channels and the resulting message aggregation mechanism, and the derivation of the resulting posterior probabilities are novel contributions of this article.

In what follows, we provide detailed descriptions of the CP and JCDE phases of the proposed technique briefly explained above.

Let us define the soft replicas of $x_m[k]$ and $h_{nm}[k]$ as $\hat{x}_{mk}$ and $\hat{h}_{k,nm}$, respectively, such that their MSEs can be respectively expressed as

$$\hat{\psi}^{\mathrm{x}}_{mk} \triangleq \mathbb{E}\left[|\tilde{x}_{mk}|^2\right], \text{ with } \tilde{x}_{mk} \triangleq x_m[k] - \hat{x}_{mk}, \tag{15a}$$

$$\hat{\psi}^{\mathrm{h}}_{k,nm} \triangleq \mathbb{E}\left[|\tilde{h}_{k,nm}|^2\right], \text{ with } \tilde{h}_{k,nm} \triangleq h_{nm}[k] - \hat{h}_{k,nm}, \tag{15b}$$

where the quantities $\tilde{x}_{mk}$ and $\tilde{h}_{k,nm}$ denote the estimation errors, respectively.

Denoting the soft replicas of $\boldsymbol{h}_m[k]$ and $\boldsymbol{H}[k]$ by $\hat{\boldsymbol{h}}_{k,m} \triangleq \left[\hat{h}_{k,1m}, \ldots, \hat{h}_{k,nm}, \ldots, \hat{h}_{k,Nm}\right]^{\mathsf{T}}$ and $\hat{\boldsymbol{H}}_k \triangleq \left[\hat{\boldsymbol{h}}_{k,1}, \ldots, \hat{\boldsymbol{h}}_{k,m}, \ldots, \hat{\boldsymbol{h}}_{k,M}\right] \in \mathbb{C}^{N\times M}$, respectively, in similarity to the above, their corresponding error covariance matrices can be expressed as

$$\hat{\boldsymbol{\Psi}}^{\mathrm{h}}_{k,m} \triangleq \mathbb{E}\left[\tilde{\boldsymbol{h}}_{k,m}\tilde{\boldsymbol{h}}^{\mathsf{H}}_{k,m}\right] \in \mathbb{C}^{N\times N}, \text{ with } \tilde{\boldsymbol{h}}_{k,m} \triangleq \boldsymbol{h}_m[k] - \hat{\boldsymbol{h}}_{k,m}, \tag{16a}$$

$$\hat{\boldsymbol{\Psi}}^{\mathrm{h}}_{k} \triangleq \mathbb{E}\left[\tilde{\boldsymbol{H}}_k\tilde{\boldsymbol{H}}^{\mathsf{H}}_{k}\right] \in \mathbb{C}^{N\times N}, \text{ with } \tilde{\boldsymbol{H}}_k \triangleq \boldsymbol{H}[k] - \hat{\boldsymbol{H}}_k. \tag{16b}$$

For simplicity, we assume that quadrature phase-shift keying (QPSK) modulation with unit power is employed, where the constellation is expressed as $\mathcal{X} = \left\{\pm\frac{1}{\sqrt{2}} \pm \mathrm{j}\frac{1}{\sqrt{2}}\right\}$.

A. CP Phase

Let us start with the CP phase, which uses (12) describing channel state transitions to generate the prior estimates (*i.e.*, initial soft replicas used in the subsequent JCDE phase) of $\boldsymbol{H}[k], \forall k \in \mathcal{K}_\tau$. Since soft replicas are not available at the first step ($\tau = 1$), the initial conditional expectation of $\boldsymbol{H}[k]$, given $\boldsymbol{H}[0]$, is calculated via (12) as[4]

$$\hat{\boldsymbol{H}}_k = \mathbb{E}_{\mathbf{H}[k]|\mathbf{H}[0]}\left[\boldsymbol{H}[k]\middle|\boldsymbol{H}[0]\right] = r^k\boldsymbol{H}[0], \tag{17}$$

where its corresponding second-order statistics can be obtained from (44) in Appendix as

$$\hat{\boldsymbol{\Psi}}^{\mathrm{h}}_{k} = \mathbb{E}_{\mathbf{H}[k]|\mathbf{H}[0]}\left[\boldsymbol{H}[k]\boldsymbol{H}[k]^{\mathsf{H}}\middle|\boldsymbol{H}[0]\right] = \boldsymbol{\Omega}_k \in \mathbb{C}^{N\times N}, \tag{18a}$$

$$\hat{\boldsymbol{\Psi}}^{\mathrm{h}}_{k,m} = \mathbb{E}_{\mathbf{h}_m[k]|\mathbf{h}_m[0]}\left[\boldsymbol{h}_m[k]\boldsymbol{h}_m[k]^{\mathsf{H}}\middle|\boldsymbol{h}_m[0]\right] = \boldsymbol{\Omega}_{k,m} \in \mathbb{C}^{N\times N}, \tag{18b}$$

$$\psi^{\mathrm{h}}_{k,nm} = \mathbb{E}_{\mathsf{h}_{nm}[k]|\mathsf{h}_{nm}[0]}\left[|h_{nm}[k]|^2\middle|h_{nm}[0]\right] = \omega_{k,nm}. \tag{18c}$$

Owing to the availability of soft replicas as prior information at the second and subsequent steps ($\tau \neq 1$), the conditional expectation of $\boldsymbol{H}[k]$ can thereafter be obtained as

$$\begin{aligned}\hat{\boldsymbol{H}}_k &= \mathbb{E}_{\mathbf{H}[k]|\mathbf{H}[0],\hat{\mathbf{H}}_k,\forall k\in\breve{\mathcal{K}}_{\tau-1}}\left[\boldsymbol{H}[k]\middle|\boldsymbol{H}[0],\hat{\boldsymbol{H}}_k,\forall k\in\breve{\mathcal{K}}_{\tau-1}\right]\\ &\approx \mathbb{E}_{\mathbf{H}[k]|\hat{\mathbf{H}}_k,\forall k\in\mathcal{K}_{\tau-1}}\left[\boldsymbol{H}[k]\middle|\hat{\boldsymbol{H}}_k,\forall k\in\mathcal{K}_{\tau-1}\right] &\text{(19a)}\\ &\approx \mathbb{E}_{\mathbf{H}[k]|\hat{\mathbf{H}}_{|\breve{\mathcal{K}}_{\tau-1}|-1}}\left[\boldsymbol{H}[k]\middle|\hat{\boldsymbol{H}}_{|\breve{\mathcal{K}}_{\tau-1}|-1}\right], &\text{(19b)}\end{aligned}$$

where for simplicity we have denoted $\breve{\mathcal{K}}_\tau \triangleq \bigcup_{i=1}^{\tau}\mathcal{K}_\tau$. Since all estimates are computed based on the previous estimates and $\boldsymbol{H}[0]$, the given knowledge is not independent of each other. Therefore, if the sequential estimation process works well based on the Bayes-optimal inference framework, it is sufficient to use as prior information only the estimates obtained in the immediately previous step, as in (19a). In addition, as will be explained later in the JCDE phase, $\hat{\boldsymbol{H}}_k, \forall k \in \mathcal{K}_{\tau-1}$, is generated using all knowledge of $\boldsymbol{y}[k], \forall k \in \mathcal{K}_{\tau-1}$, so in theory, we can create $\hat{\boldsymbol{H}}_k$ from any appropriate estimate contained in $\mathcal{K}_{\tau-1}$, *e.g.*, the latest estimated channel $\hat{\boldsymbol{H}}_{|\breve{\mathcal{K}}_{\tau-1}|-1}$, as in (19b). However, the JCDE algorithm is sub-optimal; hence, the reliability of $\hat{\boldsymbol{H}}_k$ varies randomly for each k, depending on the convergence behavior of iterative estimation, and thus the low accuracy of the selected estimate may induce serious performance degradation due to error propagation.

To alleviate this inconvenience, we employ a simple criterion to select as prior information the estimate with the minimum MSE among the estimates used in the τ-th step, as follows:

$$k_\tau = \underset{\underline{k}\in\mathcal{K}_\tau\setminus\mathcal{K}_\tau^+}{\operatorname{argmin}} \sum_{n=1}^{N}\sum_{m=1}^{M}\hat{\psi}^{\mathrm{h}}_{\underline{k},nm}, \tag{20}$$

$$\hat{\boldsymbol{H}}_k = \mathbb{E}_{\mathbf{H}[k]|\hat{\mathbf{H}}_{k_\tau}}\left[\boldsymbol{H}[k]\middle|\hat{\boldsymbol{H}}_{k_\tau}\right] = r^{k-k_\tau}\hat{\boldsymbol{H}}_{k_\tau}, \ \forall k \in \mathcal{K}_\tau^{>k_\tau}, \tag{21}$$

where $\mathcal{K}_\tau^{>k_\tau} \triangleq \{k \in \mathcal{K}_\tau | k > k_\tau\}$, and its second-order statistics can be obtained from (15b), (16), and (44) as

$$\hat{\boldsymbol{\Psi}}^{\mathrm{h}}_{k} = \mathbb{E}_{\mathbf{H}[k]|\hat{\mathbf{H}}_{k_\tau}}\left[\boldsymbol{H}[k]\boldsymbol{H}[k]^{\mathsf{H}}\middle|\hat{\boldsymbol{H}}_{k_\tau}\right]$$

[4]In order to avoid being repetitive, we hereafter omit the qualifier "$\forall k \in \mathcal{K}_\tau$" when obvious.

$$= r^{2(k-k_\tau)}\hat{\boldsymbol{\Psi}}^{\mathrm{h}}_{k_\tau} + \boldsymbol{\Omega}_{k-k_\tau}, \tag{22a}$$

$$\hat{\boldsymbol{\Psi}}^{\mathrm{h}}_{k,m} = \mathbb{E}_{\mathbf{h}_m[k]|\hat{\mathbf{h}}_{k_\tau,m}}\left[\boldsymbol{h}_m[k]\boldsymbol{h}_m[k]^{\mathsf{H}}\middle|\hat{\boldsymbol{h}}_{k_\tau,m}\right]$$
$$= r^{2(k-k_\tau)}\hat{\boldsymbol{\Psi}}^{\mathrm{h}}_{k_\tau,m} + \boldsymbol{\Omega}_{k-k_\tau,m}, \tag{22b}$$

$$\psi^{\mathrm{h}}_{k,nm} = \mathbb{E}_{\mathsf{h}_{nm}[k]|\hat{\mathsf{h}}_{k_\tau,nm}}\left[|h_{nm}[k]|^2\middle|\hat{h}_{k_\tau,nm}\right]$$
$$= r^{2(k-k_\tau)}\hat{\psi}^{\mathrm{h}}_{k_\tau,m} + \omega_{k-k_\tau,nm}. \tag{22c}$$

In the above, a single CP phase is described. For the discrete-time indices that were included in the active set $\mathcal{K}_{\tau-1}$ and were not updated in the τ-th CP phase, the soft replicas obtained at the $\tau-1$ step is used as the prior estimate as they are.

B. JCDE Phase

Herein, the JCDE algorithm for estimating $x_m[k]$ and $h_{nm}[k]$ is focused upon. Note that the following process is performed on the RX symbols $y_n[k], \forall n \in \mathcal{N} \triangleq \{1,2,\ldots,N\}, \forall k \in \mathcal{K}_\tau$, in parallel to estimate all TX symbols $x_m[k], \forall m \in \mathcal{M} \triangleq \{1,2,\ldots,M\}, \forall k \in \mathcal{K}_\tau$, and channel coefficients $h_{nm}[k], \forall (n,m), \forall k \in \mathcal{K}_\tau$, simultaneously. In the sequel, we refrain from explicitly including the qualifier "$\forall k \in \mathcal{K}_\tau$" after each equation, in order to avoid being repetitive.

1) FN Process for DD: The iterative process, which is carried out for the RX vector, starts at the FNs with soft interference cancellation (soft IC) performed on $\boldsymbol{y}[k]$ with utilizing all the soft replicas obtained in the previous JCDE iteration ($t \neq 1$). At the first iteration ($t = 1$), the soft replicas $\hat{x}_{mk}, \forall k \in \mathcal{K}_\tau$, are appropriately initialized, *e.g.*, $\hat{x}_{mk} = 0, \forall k \in \mathcal{K}_\tau$, and $\hat{h}_{k,nm}, \forall k \in \mathcal{K}_\tau$, are obtained in the CP phase. In the detection of an arbitrary TX symbol $x_m[k]$, the cancellation process is expressed as (23), shown at the bottom of the next page. The residual interference-plus-noise component of (23) can be approximated as a multivariate complex Gaussian random vector; this approximation is referred to as vector Gaussian approximation (VGA) [49]. Thus, the conditional PDF of (23), given $x_m[k]$, can be expressed as

$$p_{\tilde{\mathbf{y}}_{m,k}|\mathsf{x}_m[k]}\left(\tilde{\boldsymbol{y}}_{m,k}|x_m[k]\right)$$
$$\propto \exp\left[-\left(\tilde{\boldsymbol{y}}_{m,k} - \hat{\boldsymbol{h}}_{k,m}x_m[k]\right)^{\mathsf{H}}\boldsymbol{\Xi}^{-1}_{k,m}\left(\tilde{\boldsymbol{y}}_{m,k} - \hat{\boldsymbol{h}}_{k,m}x_m[k]\right)\right], \tag{24}$$

with

$$\boldsymbol{\Xi}_{m,k} \triangleq \mathbb{E}\left[\boldsymbol{\kappa}^{\mathrm{x}}_{k,m}\left(\boldsymbol{\kappa}^{\mathrm{x}}_{k,m}\right)^{\mathsf{H}}\right]$$
$$= \underbrace{\sum_{m=1}^{M}\hat{\psi}^{\mathrm{x}}_{mk}\hat{\boldsymbol{h}}_{k,m}\hat{\boldsymbol{h}}^{\mathsf{H}}_{k,m} + \hat{\boldsymbol{\Psi}}^{\mathrm{h}}_{k} + N_0\boldsymbol{I}_N - \hat{\psi}^{\mathrm{x}}_{mk}\hat{\boldsymbol{h}}_{k,m}\hat{\boldsymbol{h}}^{\mathsf{H}}_{k,m}}_{\triangleq\boldsymbol{\Xi}_k\in\mathbb{C}^{N\times N}}. \tag{25}$$

2) VN Process for DD: From (24), the beliefs corresponding to $x_m[k]$ are combined over all RX antenna indices via linear minimum mean square error (MMSE) filters, which results in the PDF of a joint belief l^{x}_{mk} for $x_m[k]$. This is expressed as

$$p_{\mathsf{l}^{\mathrm{x}}_{mk}|\mathsf{x}_m[k]}\left(l^{\mathrm{x}}_{mk}|x_m[k]\right) = p_{\tilde{\mathbf{y}}_{m,k}|\mathsf{x}_m[k]}\left(\tilde{\boldsymbol{y}}_{m,k}|x_m[k]\right)$$

TABLE II

EXAMPLES OF $\mathcal{S}_{k,\tau}$ FOR THE CASE OF $(W,D,G) = (2,3,6)$, WHERE $K = 8$

τ \ k	0	1	2	3	4	5	6	7
1	{1}	{0}	-	-	-	-	-	-
2	{1,2,3}	{0,2,3}	{0,1,3}	{0,1,2}	-	-	-	-
3	{1,2,3}	{0,2,3,4}	{0,1,3,4,5}	{0,1,2,4,5}	{1,2,3,5}	{2,3,4}	-	-
4	-	-	{0,1,3,4,5}	{0,1,2,4,5}	{1,2,3,5,6,7}	{2,3,4,6,7}	{3,4,5,7}	{4,5,6}
5	-	-	-	-	{1,2,3,5,6,7}	{2,3,4,6,7}	{3,4,5,7}	{4,5,6}
6	-	-	-	-	-	-	{3,4,5,7}	{4,5,6}

$$\propto \exp\left[-\frac{|x_m[k] - \overline{x}_{mk}|^2}{\overline{\psi}^{\mathrm{x}}_{mk}}\right], \tag{26}$$

where $\overline{x}_{mk} = \frac{1}{\eta_{mk}}\hat{\boldsymbol{h}}^{\mathsf{H}}_{k,m}\boldsymbol{\Xi}^{-1}_k\tilde{\boldsymbol{y}}_{m,k}$, $\overline{\psi}^{\mathrm{x}}_{mk} = \frac{1-\eta_{mk}\hat{\psi}^{\mathrm{x}}_{mk}}{\eta_{mk}}$, and $\eta_{mk} = \hat{\boldsymbol{h}}^{\mathsf{H}}_m\boldsymbol{\Xi}^{-1}_k\hat{\boldsymbol{h}}_m$, which are derived by applying the matrix inversion lemma to eliminate the dependency of $\boldsymbol{\Xi}_{m,k}$ in (24) on the index $m \in \mathcal{M}$, enabling us to share the same inverse matrix $\boldsymbol{\Xi}^{-1}_k$ for $\forall m$ [50].

Assuming that the effective noise components in $l^{\mathrm{x}}_{mk}, \forall (m,k)$, are not correlated to each other under VGA, using Bayes' rule, the soft replica of $x_m[k]$ and its MSE can be in general obtained from the symbol-wise conditional expectation, given l^{x}_{mk}, as

$$\hat{x}_{mk} = \sum_{\chi\in\mathcal{X}}\frac{\chi\cdot p_{\mathsf{l}^{\mathrm{x}}_{mk}|\mathsf{x}_m[k]}\left(l^{\mathrm{x}}_{mk}|\chi\right)p_{\mathsf{x}_m[k]}\left(\chi\right)}{\sum\limits_{\chi'\in\mathcal{X}}p_{\mathsf{l}^{\mathrm{x}}_{mk}|\mathsf{x}_m[k]}\left(l^{\mathrm{x}}_{mk}|\chi'\right)p_{\mathsf{x}_m[k]}\left(\chi'\right)}, \tag{27a}$$

$$\hat{\psi}^{\mathrm{x}}_{mk} = \sum_{\chi\in\mathcal{X}}\frac{|\chi-\hat{x}_{mk}|^2\cdot p_{\mathsf{l}^{\mathrm{x}}_{mk}|\mathsf{x}_m[k]}\left(l^{\mathrm{x}}_{mk}|\chi\right)p_{\mathsf{x}_m[k]}\left(\chi\right)}{\sum\limits_{\chi'\in\mathcal{X}}p_{\mathsf{l}^{\mathrm{x}}_{mk}|\mathsf{x}_m[k]}\left(l^{\mathrm{x}}_{mk}|\chi'\right)p_{\mathsf{x}_m[k]}\left(\chi'\right)}, \tag{27b}$$

where the denominator in the integrand is introduced for normalization purposes.

In Gray-coded QPSK signaling with unit power, (27) can be readily obtained by the following *denoiser* as [51]

$$\hat{x}_{mk} = \frac{1}{\sqrt{2}}\cdot\left(\tanh\left[\frac{\sqrt{2}\Re\{\overline{x}_{mk}\}}{\overline{\psi}^{\mathrm{x}}_{mk}}\right] + \mathrm{j}\tanh\left[\frac{\sqrt{2}\Im\{\overline{x}_{mk}\}}{\overline{\psi}^{\mathrm{x}}_{mk}}\right]\right), \tag{28a}$$

$$\hat{\psi}^{\mathrm{x}}_{mk} = 1 - |\hat{x}_{mk}|^2. \tag{28b}$$

3) FN Process for CE: In the CE process, it is necessary to perform belief combining across the discrete-time dimension; however, since the channels vary with respect to k, the conventional belief consensus mechanism does not work well, which results in significant loss of diversity gain. Therefore, it is essential to perform state transitions using (12) in the belief domain as well. Furthermore, it is effective to restrict the beliefs to be combined to those generated based on discrete-time observations close to the index k to be estimated.

In this article, a set of discrete-time indices of the beliefs to be combined when estimating the entries of $\boldsymbol{H}[k]$ at the τ-th step is given by the vicinity of k as

$$\mathcal{S}_{k,\tau} \triangleq \left\{s \in \breve{\mathcal{K}}_\tau \setminus \{k\}\middle| k - \frac{G}{2} \leq k \leq k + \frac{G}{2}\right\}, \tag{29}$$

where G is the number of beliefs to be combined and should be even. Table II lists the sets $\mathcal{S}_{k,\tau}$ for each k when we set to $(W,D,G) = (2,3,6)$, with $K = 8$, as an example.

In what follows, we provide detailed descriptions of the estimation of an arbitrary $h_{nm}[k]$. Let us start with the soft IC on $y_n[s], \forall s \in \mathcal{S}_{k,\tau}$, with the aid of soft replicas, yields (30), shown at the bottom of the next page. Here, we note that the statistical property of beliefs propagated from past discrete times and those propagated from future discrete times are different, so that we apply state transitions to the channel coefficients in (30) to extract the information of $h_{nm}[k]$. For the beliefs with the discrete-time index $s < k$, using (12), (30) can be rewritten as

$$\tilde{y}_{s,nm} = \left(r^{s-k} h_{nm}[k] - r^{s-k} \zeta^{\mathrm{h}}_{s\to k,nm}\right) \hat{x}_{ms} + \kappa^{\mathrm{h}}_{s,nm}, \tag{31a}$$

$$r^{k-s} \tilde{y}_{s,nm} = h_{nm}[k] \hat{x}_{ms} - \zeta^{\mathrm{h}}_{s\to k,nm} \hat{x}_{ms} + r^{k-s} \kappa^{\mathrm{h}}_{s,nm}, \tag{31b}$$

where we define the errors due to channel aging as

$$\zeta^{\mathrm{h}}_{s\to k,nm} \triangleq \sqrt{\frac{1-r^2}{L}} \sum_{l=1}^{L} \sum_{c=1}^{C_\ell} \frac{[\boldsymbol{A}_{\ell,c}]_{n,m}}{\sqrt{C_\ell}} \beta_{\ell,c}(k, k-s). \tag{32}$$

Assuming that the scalar Gaussian approximation (SGA) [49] can be applied, namely, that the effective noise component of (31b) – which includes errors due to channel aging, residual interference, and AWGN – can be approximated as a complex scalar Gaussian random variable, the conditional PDF of (31b), given $h_{nm}[k]$, can be expressed with (33), shown at the bottom of the next page, as

$$p_{r^{k-s}\tilde{\mathrm{y}}_{s,nm}|\mathrm{h}_{nm}[k]}\left(r^{k-s}\tilde{y}_{s,nm}|h_{nm}[k]\right) \propto \exp\left[-\frac{\left|r^{k-s}\tilde{y}_{s,nm} - h_{nm}[k]\hat{x}_{ms}\right|^2}{\nu^{\mathrm{h}}_{s\to k,nm}}\right], \tag{34}$$

where the instantaneous channel gain $|h_{nm}[s]|^2$ is not available, and therefore is here approximated by the true variance, *i.e.*,

$$\theta_{nm} \triangleq \mathbb{E}\left[|h_{nm}[s]|^2\right] = \frac{1}{L} \sum_{l=1}^{L} \sum_{c=1}^{C_\ell} \frac{\left|[\boldsymbol{A}_{\ell,c}]_{n,m}\right|^2}{C_\ell}. \tag{35}$$

In a similar manner, for the beliefs with $s > k$, (31) can be rewritten as

$$\tilde{y}_{s,nm} = \Bigg(r^{s-k} h_{nm}[k] + \underbrace{\sqrt{\frac{1-r^2}{L}} \sum_{l=1}^{L} \sum_{c=1}^{C_\ell} \frac{[\boldsymbol{A}_{\ell,c}]_{nm}}{\sqrt{C_\ell}} \beta_{\ell,c}(s, s-k)}_{=\zeta^{\mathrm{h}}_{s\to k,nm}}\Bigg) \hat{x}_{ms} + \kappa^{\mathrm{h}}_{s,nm}, \tag{36a}$$

$$r^{k-s} \tilde{y}_{s,nm} = h_{nm}[k] \hat{x}_{ms} + r^{k-s} \zeta^{\mathrm{h}}_{s\to k,nm} \hat{x}_{ms} + r^{k-s} \kappa^{\mathrm{h}}_{s,nm}. \tag{36b}$$

It follows that the conditional PDF of (36b), given $h_{nm}[k]$, can be expressed as

$$p_{r^{k-s}\tilde{\mathrm{y}}_{s,nm}|\mathrm{h}_{nm}[k]}\left(r^{k-s}\tilde{y}_{s,nm}|h_{nm}[k]\right) \propto \exp\left[-\frac{\left|r^{k-s}\tilde{y}_{s,nm} - h_{nm}[k]\hat{x}_{ms}\right|^2}{\nu^{\mathrm{h}}_{s\to k,nm}}\right], \tag{37}$$

with

$$\nu^{\mathrm{h}}_{s\to k,nm} = \mathbb{E}\left\{\left|r^{k-s}\zeta^{\mathrm{h}}_{s\to k,nm}\hat{x}_{ms} + r^{k-s}\kappa^{\mathrm{h}}_{s,nm}\right|^2\right\} = r^{2(k-s)}\left(\omega_{s-k,nm}\left|\hat{x}_{ms}\right|^2 + \nu^{\mathrm{h}}_{s,nm}\right). \tag{38}$$

4) VN Process for CE: Again, assuming that a high-precision SGA of the effective noise components among $\tilde{y}_{s,nm}, \forall s \in \mathcal{S}_{k,\tau}$, can be used, the beliefs corresponding to $h_{nm}[k]$ are combined over $\forall s \in \mathcal{S}_{k,\tau}$, which results in the PDF of an *extrinsic* belief $l^{\mathrm{h}}_{k,nm}$, given $h_{nm}[k]$, yielding

$$p_{l^{\mathrm{h}}_{k,nm}|\mathrm{h}_{nm}[k]}\left(l^{\mathrm{h}}_{k,nm}|h_{nm}[k]\right) = \prod_{s\in\mathcal{S}_{k,\tau}} p_{r^{k-s}\tilde{\mathrm{y}}_{s,nm}|\mathrm{h}_{nm}[k]}\left(r^{k-s}\tilde{y}_{s,nm}|h_{nm}[k]\right) \propto \exp\left[-\frac{\left|h_{nm}[k] - \overline{h}_{k,nm}\right|^2}{\overline{\psi}^{\mathrm{h}}_{k,nm}}\right], \tag{39}$$

with

$$\overline{h}_{k,nm} = \overline{\psi}^{\mathrm{h}}_{k,nm} \sum_{s\in\mathcal{S}_{k,\tau}} \frac{\hat{x}^*_{mi} r^{k-s} \tilde{y}_{s,nm}}{\nu^{\mathrm{h}}_{s\to k,nm}}, \tag{40a}$$

$$\overline{\psi}^{\mathrm{h}}_{k,nm} = \left(\sum_{s\in\mathcal{S}_{k,\tau}} \frac{\left|\hat{x}_{ms}\right|^2}{\nu^{\mathrm{h}}_{s\to k,nm}}\right)^{-1}, \tag{40b}$$

where we define $\overline{\boldsymbol{h}}_{k,m} \triangleq \left[\overline{h}_{k,1m}, \ldots, \overline{h}_{k,Nm}\right]^{\mathsf{T}}$ and $\overline{\boldsymbol{\Psi}}^{\mathrm{h}}_{k,m} \triangleq \mathrm{diag}\left[\overline{\psi}^{\mathrm{h}}_{k,1m}, \ldots, \overline{\psi}^{\mathrm{h}}_{k,Nm}\right]$, respectively.

The extrinsic combining operation in (39) enables significant suppression of the harmful effect of self-feedback from the previous iteration, by removing the beliefs of $\tilde{y}_{k,nm}$ from $l^{\mathrm{h}}_{n,mk}$. Assuming that the effective noise components in $l^{\mathrm{h}}_{k,nm}, \forall n$, are not correlated to each other under the SGA, the soft replica of $\boldsymbol{h}_m[k]$ can be in general obtained from the vector-wise conditional expectation, given $l^{\mathrm{h}}_{k,nm}, \forall n$, as

$$\hat{\boldsymbol{h}}_{k,m} = \frac{\int_{\boldsymbol{h}_m[k]} \boldsymbol{h}_m[k] \prod_{i=1}^{N} p_{l^{\mathrm{h}}_{k,im}|\mathrm{h}_{im}[k]}\left(l^{\mathrm{h}}_{k,im}|h_{im}[k]\right) p_{\mathbf{h}_m[k]}\left(\boldsymbol{h}_m[k]\right)}{\int_{\boldsymbol{h}'_m[k]} \prod_{i=1}^{N} p_{l^{\mathrm{h}}_{k,im}|\mathrm{h}_{im}[k]}\left(l^{\mathrm{h}}_{k,im}|h'_{im}[k]\right) p_{\mathbf{h}_m[k]}\left(\boldsymbol{h}'_m[k]\right)}. \tag{41}$$

$$\tilde{\boldsymbol{y}}_{m,k} = \boldsymbol{y}[k] - \underbrace{\sum_{i\neq m}^{M} \hat{\boldsymbol{h}}_{k,i}\hat{x}_{ik}}_{\text{Inter-symbol interference cancellation}} = \hat{\boldsymbol{h}}_{k,m}x_m[k] + \underbrace{\tilde{\boldsymbol{h}}_{k,m}x_m[k] + \sum_{i\neq m}^{M}\left(\boldsymbol{h}_i[k]x_i[k] - \hat{\boldsymbol{h}}_{k,i}\hat{x}_{ik}\right) + \boldsymbol{z}[k]}_{\triangleq \boldsymbol{\kappa}^{\mathrm{x}}_{k,m}:\text{Residual interference plus noise}} \tag{23}$$

Using the multiplication rule [29] in (41), with $\boldsymbol{h}_m[k] \sim \mathcal{CN}(r^k \boldsymbol{h}_m[0], \boldsymbol{\Omega}_{k,m}), \forall k$, yields

$$\hat{\boldsymbol{h}}_{k,m} = C \int_{\boldsymbol{h}_m[k]} \boldsymbol{h}_m[k]\ \mathcal{CN}\Big(\boldsymbol{h}_m[k]; \boldsymbol{\Omega}_{k,m}\boldsymbol{\Lambda}_{k,m}^{-1}\overline{\boldsymbol{h}}_{k,m} + r^k \overline{\boldsymbol{\Psi}}^{\mathrm{h}}_{k,m}\boldsymbol{\Lambda}_{k,m}^{-1}\boldsymbol{h}_m[0]\,, \boldsymbol{\Omega}_{k,m}\boldsymbol{\Lambda}_{k,m}^{-1}\overline{\boldsymbol{\Psi}}^{\mathrm{h}}_{k,m}\Big), \tag{42}$$

where $\boldsymbol{\Lambda}_{k,m} \triangleq \boldsymbol{\Omega}_{k,m} + \overline{\boldsymbol{\Psi}}^{\mathrm{h}}_{k,m} \in \mathbb{C}^{N\times N}$, and C is Pearl's normalization constant. From (42), the soft replica of $\boldsymbol{h}_m[k]$ and its MSE are respectively expressed as

$$\hat{\boldsymbol{h}}_{k,m} = \boldsymbol{\Omega}_{k,m}\boldsymbol{\Lambda}_{k,m}^{-1}\overline{\boldsymbol{h}}_{k,m} + r^k \overline{\boldsymbol{\Psi}}^{\mathrm{h}}_{k,m}\boldsymbol{\Lambda}_{k,m}^{-1}\boldsymbol{h}_m[0], \tag{43a}$$

$$\hat{\boldsymbol{\Psi}}^{\mathrm{h}}_{k,m} = \boldsymbol{\Omega}_{k,m}\boldsymbol{\Lambda}_{k,m}^{-1}\overline{\boldsymbol{\Psi}}^{\mathrm{h}}_{k,m}. \tag{43b}$$

When the number of iterations reaches the predetermined value, one JCDE phase is completed.

C. Algorithm Description

The pseudo-code of the proposed JCTDD algorithm described above is given in Algorithms 1–3. Besides the RX signal vectors $\boldsymbol{y}[k], \forall k$, and the estimated initial effective channel matrix $\boldsymbol{H}[0]$, the algorithm requires only system parameters such as $\boldsymbol{A}_{\ell,c}, \forall\,(l,c)$, the correlation parameter r, the noise spectral density N_0, the predetermined parameters $\{W, D, G\}$, and the maximum number of iterations in the JCDE phase $t_{\max}$, outputting the hard-decision estimates of the TX symbols $\check{x}_m[k], \forall(k,m)$, and the estimates of the channels $\check{\boldsymbol{H}}[k], \forall k$. The proposed JCTDD algorithm tracks the channel in order to aid the DD, indicating that it can be seen as an advanced DD algorithm. Besides, the inputs required for the proposed algorithm are the same as the conventional linear MMSE receiver. No additional control signaling is required to perform the proposed method; therefore, to the best of our knowledge, there is no increase in signaling overhead due to the proposed receiver and no changes to higher-level functions or protocols are required. Notice that lines 7 and 23 in Algorithm 3 correspond to a well-known belief damping procedure [33], [49], which aims to avoid the iterative estimation being trapped at a local optimum, especially at the early stage of the iterations by allowing an inertial update of the quantities $\hat{x}_{mk}$, $\hat{\psi}^{\mathrm{x}}_{mk}$, $\hat{\boldsymbol{h}}_{k,m}$, and $\hat{\boldsymbol{\Psi}}^{\mathrm{h}}_{k,m}$. This is due to the fact that at the early stage of the iterative process, the Gaussian approximation assumed in (24), (34), and (37) may not capture the actual statistics of the effective noise, which might lead to convergence to a local optimum point [52].

Even if the communication environment changes, the proposed algorithm can be applied as-is if the input covariance matrices contain appropriate statistical information. For instance, it can be implemented as-is regardless of the operating frequency band (although further algorithmic improvements can be made in case of sub-6GHz bands since the current proposed method is tailored for mmWave bands).[5]

One may have already noticed, the proposed algorithm can be interpreted as a natural extension of the KF mechanism to the BBI framework, while approximating the channel aging interference component as white Gaussian noise. However, a method that repeats prediction and update phases every discrete time step (*i.e.*, $W = D = 1$), as in the classical KF, cannot achieve highly accurate CE/CT, which requires diversity combining in the discrete-time dimension. The proposed method solves the latter problem by employing a "window-sliding" strategy in which each step is processed in time blocks of size $W \times D$. In addition, by setting $D > 1$, the scheme takes advantage of past estimates, mitigating the inherent instability of initial convergence of the JCDE algorithm. It is also worth mentioning that the harmful effect of channel aging caused by the extension to block-wise processing is suppressed by incorporating the transition model into the belief update rules in (31) and (36), and as a final remark, we emphasize that although we derived the JCTDD algorithm as a pure DD scheme based on the knowledge of the estimated channel $\boldsymbol{H}[0]$ in this article, the approach can also be directly applied to systems that track time variations in the channel by transmitting pilot symbols with short periods. By assigning the pilot symbols to $\hat{x}_{mk}, \forall(k \in \mathcal{K}_{\mathrm{p}}, m)$, and 0 to the corresponding

[5]Similarly, the proposed algorithm can be applied to multi-user scenarios. Of course, large-fading factors (*e.g.* path-loss) and TX power for each user must be estimated in the beam training step or earlier preprocessing. However, if these effects are embedded in the effective channel matrix $\boldsymbol{H}_f[k]$ in (11) and reflected in the corresponding covariance matrices, there is no need to change the structure of the algorithm. Note that the time correlation parameter r becomes a variable that depends on the user index, some trivial mathematical manipulations must be added in the message update formula accordingly.

$$\tilde{y}_{s,nm} = y_n[s] - \underbrace{\sum_{i\neq m}^{M} \hat{h}_{s,ni}\hat{x}_{is}}_{\text{Inter-symbol interference cancellation}} = h_{nm}[s]\hat{x}_{ms} + \underbrace{h_{nm}[s]\tilde{x}_{ms} + \sum_{i\neq m}^{M}\Big(h_{ni}[s]x_i[s] - \hat{h}_{s,ni}\hat{x}_{is}\Big) + z_n[s]}_{\triangleq \kappa^{\mathrm{h}}_{s,nm}\text{: Residual interference plus noise}} \tag{30}$$

$$\begin{aligned}\nu^{\mathrm{h}}_{s\to k,nm} &= \mathbb{E}\left[\left|-\zeta^{\mathrm{h}}_{s\to k,nm}\hat{x}_{ms} + r^{k-s}\kappa^{\mathrm{h}}_{s,nm}\right|^2\right] \\ &= \omega_{k-s,nm}\left|\hat{x}_{ms}\right|^2 + r^{2(k-s)}\underbrace{\left[\sum_{i\neq m}^{M}\left\{\left|\hat{h}_{s,ni}\right|^2\hat{\psi}^{\mathrm{x}}_{is} + \left(\left|\hat{x}_{is}\right|^2 + \hat{\psi}^{\mathrm{x}}_{is}\right)\hat{\psi}^{\mathrm{h}}_{s,ni}\right\} + \theta_{nm}\hat{\psi}^{\mathrm{x}}_{ms} + N_0\right]}_{\triangleq \nu^{\mathrm{h}}_{s,nm}},\end{aligned} \tag{33}$$

Algorithm 1 - BBI-Based JCTDD Algorithm for Time-Varying mmWave MIMO Systems

Input: $\{\boldsymbol{y}[k], \forall k\}$, $\boldsymbol{H}[0]$, $\{\boldsymbol{A}_{\ell,c}, \forall(\ell,c)\}$, r, N_0, W, D, G, $t_{\max}$
Output: $\{\check{x}_m[k], \forall(k,m)\}$, $\left\{\check{\boldsymbol{H}}[k], \forall k\right\}$

/* ———— Preprocessing ———— */
Obtain using Eq. (14) $\mathcal{K}_\tau$, and generate $\mathcal{K}_\tau^+$ and $\check{\mathcal{K}}_\tau$
Obtain using Eq. (29) $\mathcal{S}_{k,\tau}$
$\forall(k,m,n)$: $\hat{x}_{mk} \leftarrow 0$ and $\psi^{\mathrm{x}}_{mk} \leftarrow 1$
$\forall(m,n)$: $\theta_{nm} \leftarrow \frac{1}{L}\sum_{l=1}^{L}\sum_{c=1}^{C_\ell}\frac{|[\boldsymbol{A}_{\ell,c}]_{n,m}|^2}{C_\ell}$
$\forall(k,m,n)$: $\omega_{k,nm} \leftarrow \left(1-r^{2k}\right)\theta_{n,m}$
$\forall m$: $\boldsymbol{\Theta}_m \leftarrow \frac{1}{L}\sum_{l=1}^{L}\sum_{c=1}^{C_\ell}\frac{[\boldsymbol{A}_{\ell,c}]_{:,m}[\boldsymbol{A}_{\ell,c}]_{:,m}^{\mathsf{H}}}{C_\ell}$
$\forall(k,m)$: $\boldsymbol{\Omega}_{k,m} \leftarrow \left(1-r^{2k}\right)\boldsymbol{\Theta}_m$
$\forall k$: $\boldsymbol{\Omega}_k \leftarrow \sum_{m=1}^{M}\boldsymbol{\Omega}_{k,m}$
/* ———— Estimation ———— */
for $\tau = 1, \ldots, \tau_{\max}$
Executing Algorithm 2 - CP Phase
Executing Algorithm 3 - JCDE Phase
end for
/* ———— Output Generation ———— */
$\forall(k,m)$: $\check{x}_m[k] \leftarrow \underset{\chi\in\mathcal{X}}{\operatorname{argmin}}\left|\chi - \hat{x}_{mk}\right|$
$\forall k$: $\check{\boldsymbol{H}}[k] \leftarrow \hat{\boldsymbol{H}}_k$

Algorithm 2 - CP Phase

Input: $\{\boldsymbol{\Omega}_k, \forall k\}$, $\{\boldsymbol{\Omega}_{k,m}, \forall(k,m)\}$, $\{\omega_{k,nm}, \forall(k,m,n)\}$, $\left\{\hat{\psi}^{\mathrm{h}}_{k,nm}, \forall(k\in\mathcal{K}_\tau, m, n)\right\}$
Output: $\left\{\hat{\boldsymbol{H}}_k, \hat{\boldsymbol{\Psi}}^{\mathrm{h}}_k, \forall k\in\mathcal{K}_\tau\right\}$, $\left\{\hat{\boldsymbol{\Psi}}^{\mathrm{h}}_{k,m}, \forall(k\in\mathcal{K}_\tau, m)\right\}$, $\left\{\hat{\psi}^{\mathrm{h}}_{k,nm}, \forall(k\in\mathcal{K}_\tau, m, n)\right\}$

if $\tau = 1$ **then**
$\forall(k\in\mathcal{K}_\tau)$: $\hat{\boldsymbol{H}}_k \leftarrow r^k\boldsymbol{H}[0]$ and $\hat{\boldsymbol{\Psi}}^{\mathrm{h}}_k \leftarrow \boldsymbol{\Omega}_k$
$\forall(k\in\mathcal{K}_\tau, m)$: $\hat{\boldsymbol{\Psi}}^{\mathrm{h}}_{k,m} \leftarrow \boldsymbol{\Omega}_{k,m}$
$\forall(k\in\mathcal{K}_\tau, m, n)$: $\hat{\psi}^{\mathrm{h}}_{k,nm} \leftarrow \omega_{k,nm}$
else if
$k_\tau \leftarrow \underset{\underline{k}\in\mathcal{K}_\tau\setminus\mathcal{K}_\tau^+}{\operatorname{argmin}}\sum_{n=1}^{N}\sum_{m=1}^{M}\hat{\psi}^{\mathrm{h}}_{\underline{k},nm}$, and generate $\mathcal{K}_\tau^{>k_\tau}$
$\forall(k\in\mathcal{K}_\tau^{>k_\tau})$: $\hat{\boldsymbol{H}}_k \leftarrow r^{k-k_\tau}\hat{\boldsymbol{H}}_{k_\tau}$ and $\hat{\boldsymbol{\Psi}}^{\mathrm{h}}_k \leftarrow r^{2(k-k_\tau)}\hat{\boldsymbol{\Psi}}^{\mathrm{h}}_{k_\tau} + \boldsymbol{\Omega}_{k-k_\tau}$
$\forall(k\in\mathcal{K}_\tau^{>k_\tau}, m)$: $\hat{\boldsymbol{\Psi}}^{\mathrm{h}}_{k,m} \leftarrow r^{2(k-k_\tau)}\hat{\boldsymbol{\Psi}}^{\mathrm{h}}_{k_\tau,m} + \boldsymbol{\Omega}_{k-k_\tau,m}$
$\forall(k\in\mathcal{K}_\tau^{>k_\tau}, m, n)$: $\hat{\psi}^{\mathrm{h}}_{k,nm} \leftarrow r^{2(k-k_\tau)}\hat{\psi}^{\mathrm{h}}_{k_\tau,m} + \omega_{k-k_\tau,nm}$
end if

MSE $\hat{\psi}^{\mathrm{x}}_{mk}, \forall(k\in\mathcal{K}_{\mathrm{p}}, m)$, where $\mathcal{K}_{\mathrm{p}}$ is the set of discrete-time indices in which the pilot symbols are embedded, the proposed method achieves Bayesian-optimal inference with the pilot symbols as addition prior information.

It shall be shown in the sequel, however, that the proposed method is already effective in estimating time-varying channels with high accuracy without the addition of pilot signals.

Algorithm 3 - JCDE Phase

Input: $\{\boldsymbol{y}[k], k\in\mathcal{K}_\tau\}$, N_0, $t_{\max}$, $\{\theta_{n,m}, \forall(m,n)\}$, $\left\{\hat{x}_{mk}, \hat{\psi}^{\mathrm{x}}_{mk}, \forall(k\in\mathcal{K}_\tau, m)\right\}$, $\left\{\hat{\boldsymbol{H}}_k, \hat{\boldsymbol{\Psi}}^{\mathrm{h}}_k, \forall k\in\mathcal{K}_\tau\right\}$, $\left\{\hat{\boldsymbol{\Psi}}^{\mathrm{h}}_{k,m}, \forall(k\in\mathcal{K}_\tau, m)\right\}$, $\left\{\hat{\psi}^{\mathrm{h}}_{k,nm}, \forall(k\in\mathcal{K}_\tau, m, n)\right\}$
Output: $\{\hat{x}_{mk}, \forall(k\in\mathcal{K}_\tau, m)\}$, $\left\{\hat{\boldsymbol{H}}_k, \forall k\in\mathcal{K}_\tau\right\}$

/* ———— JCDE Phase for DD: $\forall k\in\mathcal{K}_\tau$ ———— */
for $t = 1, \ldots, t_{\max}$
$\boldsymbol{\Xi}_k \leftarrow \sum_{m=1}^{M}\hat{\psi}^{\mathrm{x}}_{mk}\hat{\boldsymbol{h}}_{k,m}\hat{\boldsymbol{h}}^{\mathsf{H}}_{k,m} + \hat{\boldsymbol{\Psi}}^{\mathrm{h}}_k + N_0\boldsymbol{I}_N$
$\forall m$: $\eta_{mk} \leftarrow \hat{\boldsymbol{h}}^{\mathsf{H}}_m\boldsymbol{\Xi}^{-1}_k\hat{\boldsymbol{h}}_m$
$\forall m$: $\tilde{\boldsymbol{y}}_{m,k} \leftarrow \boldsymbol{y}[k] - \sum_{i\neq m}^{M}\hat{\boldsymbol{h}}_{k,i}\hat{x}_{ik}$
$\forall m$: $\overline{x}_{mk} \leftarrow \frac{1}{\eta_{mk}}\hat{\boldsymbol{h}}^{\mathsf{H}}_{k,m}\boldsymbol{\Xi}^{-1}_k\tilde{\boldsymbol{y}}_{m,k}$ and $\overline{\psi}^{\mathrm{x}}_{mk} \leftarrow \frac{1-\eta_{mk}\sigma^x_{mk}}{\eta_{mk}}$
$\forall m$:
$$\hat{x}'_{mk} \leftarrow \frac{1}{\sqrt{2}}\cdot\left(\tanh\left[\frac{\sqrt{2}\Re\{\overline{x}_{mk}\}}{\overline{\psi}^{\mathrm{x}}_{mk}}\right] + \mathrm{j}\tanh\left[\frac{\sqrt{2}\Im\{\overline{x}_{mk}\}}{\overline{\psi}^{\mathrm{x}}_{mk}}\right]\right)$$
and $\hat{\psi}'^{\mathrm{x}}_{mk} \leftarrow 1 - |\hat{x}_{mk}|^2$
$\forall m$: $\hat{x}_{mk} \leftarrow \alpha\hat{x}'_{mk} + (1-\alpha)\hat{x}_{mk}$ and $\hat{\psi}^{\mathrm{x}}_{mk} \leftarrow \alpha\hat{\psi}'^{\mathrm{x}}_{mk} + (1-\alpha)\hat{\psi}^{\mathrm{x}}_{mk}$
/* ———— JCDE Phase for CE: $\forall k\in\mathcal{K}_\tau$ ———— */
$\forall(m,n)$: $\tilde{y}_{k,nm} \leftarrow y_{nk} - \sum_{i\neq m}^{M}\hat{h}_{k,ni}\hat{x}_{ik}$
$\forall(m,n)$: $\nu^{\mathrm{h}}_{k,nm} \leftarrow \sum_{i\neq m}^{M}\left\{\left|\hat{h}_{k,ni}\right|^2\hat{\psi}^{\mathrm{x}}_{ik} + \left(|\hat{x}_{ik}|^2 + \hat{\psi}^{\mathrm{x}}_{ik}\right)\hat{\psi}^{\mathrm{h}}_{k,ni}\right\} + \theta_{nm}\hat{\psi}^{\mathrm{x}}_{mk} + N_0$
if $k > s \in \mathcal{S}_{\tau,k}$ **then**
$\forall(s\in\mathcal{S}_{\tau,k}, m, n)$: $\nu^{\mathrm{h}}_{s\to k,nm} \leftarrow \omega_{k-s,nm}|\hat{x}_{ms}|^2 + r^{2(k-s)}\nu^{\mathrm{h}}_{s,nm}$
else if $k < s \in \mathcal{S}_{\tau,k}$ **then**
$\forall(s\in\mathcal{S}_{\tau,k}, m, n)$: $\nu^{\mathrm{h}}_{s\to k,nm} \leftarrow r^{2(k-s)}\left(\omega_{s-k,nm}|\hat{x}_{ms}|^2 + \nu^{\mathrm{h}}_{s,nm}\right)$
end if
if $t = t_{\max}$ **then**
$\forall(m,n)$: $\overline{h}_{k,nm} \leftarrow \overline{\psi}^{\mathrm{h}}_{k,nm}\sum_{s\in\mathcal{S}_{k,\tau}\cup\{k\}}\frac{\hat{x}^*_{mi}r^{k-s}\tilde{y}_{s,nm}}{\nu^{\mathrm{h}}_{s\to k,nm}}$ and $\overline{\psi}^{\mathrm{h}}_{k,nm} \leftarrow \left(\sum_{s\in\mathcal{S}_{k,\tau}\cup\{k\}}\frac{|\hat{x}_{ms}|^2}{\nu^{\mathrm{h}}_{s\to k,nm}}\right)^{-1}$
else if
$\forall(m,n)$: $\overline{h}_{k,nm} \leftarrow \overline{\psi}^{\mathrm{h}}_{k,nm}\sum_{s\in\mathcal{S}_{k,\tau}}\frac{\hat{x}^*_{mi}r^{k-s}\tilde{y}_{s,nm}}{\nu^{\mathrm{h}}_{s\to k,nm}}$ and $\overline{\psi}^{\mathrm{h}}_{k,nm} \leftarrow \left(\sum_{s\in\mathcal{S}_{k,\tau}}\frac{|\hat{x}_{ms}|^2}{\nu^{\mathrm{h}}_{s\to k,nm}}\right)^{-1}$
end if
$\forall m$: $\overline{\boldsymbol{h}}_{k,m} \leftarrow \left[\overline{h}_{k,1m}, \ldots, \overline{h}_{k,Nm}\right]^{\mathsf{T}}$ and $\overline{\boldsymbol{\Psi}}^{\mathrm{h}}_{k,m} \leftarrow \mathrm{diag}\left[\overline{\psi}^{\mathrm{h}}_{k,1m}, \ldots, \overline{\psi}^{\mathrm{h}}_{k,Nm}\right]$
$\forall m$: $\boldsymbol{\Lambda}_{k,m} \leftarrow \boldsymbol{\Omega}_{k,m} + \overline{\boldsymbol{\Psi}}^{\mathrm{h}}_{k,m}$
$\forall m$: $\hat{\boldsymbol{h}}'_{k,m} \leftarrow \boldsymbol{\Omega}_{k,m}\boldsymbol{\Lambda}^{-1}_{k,m}\overline{\boldsymbol{h}}_{k,m} + r^k\overline{\boldsymbol{\Psi}}^{\mathrm{h}}_{k,m}\boldsymbol{\Lambda}^{-1}_{k,m}\boldsymbol{h}_m[0]$ and $\hat{\boldsymbol{\Psi}}'^{\mathrm{h}}_{k,m} \leftarrow \boldsymbol{\Omega}_{k,m}\boldsymbol{\Lambda}^{-1}_{k,m}\overline{\boldsymbol{\Psi}}^{\mathrm{h}}_{k,m}$
$\forall m$: $\hat{\boldsymbol{h}}_{k,m} \leftarrow \alpha\hat{\boldsymbol{h}}'_{k,m} + (1-\alpha)\hat{\boldsymbol{h}}_{k,m}$ and $\hat{\boldsymbol{\Psi}}^{\mathrm{h}}_{k,m} \leftarrow \alpha\hat{\boldsymbol{\Psi}}'^{\mathrm{h}}_{k,m} + (1-\alpha)\hat{\boldsymbol{\Psi}}^{\mathrm{h}}_{k,m}$
end for

D. Remarks on Complexity

In this subsection, let us offer a brief analysis of the computational complexities of the algorithms presented in the previous subsection. Starting with the CP phase presented

TABLE III
TIME-VARYING CHANNEL PARAMETERS BASED ON THE SPECIFICATION OF [45]

Relative speed v [kmph]	Coherence time T_c [ms]	Time correlation parameter r
15	0.507248838936000	0.999668722580203
30	0.253624419468000	0.999337554905135
45	0.169082946312000	0.999006022041674
60	0.126812209734000	0.998675548643774

in Algorithm 2, the main operations are matrix and scalar multiplications, their additions, and search over $W(D-1)$ elements. With the relationship $|\mathcal{K}_\tau^{\geq k_\tau}| < |\mathcal{K}_\tau| \leq WD$, the associated complexity is governed by line 8 in Algorithm 2, which is bounded on the order of $\mathcal{O}(N^2MWD)$.

Next, in the JCDE phase presented in Algorithm 3, the complexity of DD part is dominated by the inverse operation of $N \times N$ matrix $\boldsymbol{\Xi}_k$ and matrix and vector multiplications in line 3, which is bounded by the order of $\mathcal{O}((N^3+N^2M)WD)$. On the other hand, the complexity of CE part is dominated by M times inverse operation of $N \times N$ matrix $\boldsymbol{\Lambda}_{k,m}, \forall m$, and matrix multiplications in line 22, which is bounded on the order of $\mathcal{O}(N^3MWD)$. Other operations in the JCDE phase consist of only scalar-by-scalar calculations, whose number of four arithmetic operations is proportional to NM, as shown in [33]. In addition, the computations for the soft replica from the hyperbolic tangent function in line 6 can be handled through a look-up table, which is in order of $\mathcal{O}(MWD)$. Thus, considering the maximum number of iterations $t_{\max}$, the computational complexity of one JCDE phase is in order of $\mathcal{O}(t_{\max}N^3MWD)$.

It follows from the above that the complexity of Algorithms 1–3 is dominated by the JCDE phase, which is bounded on the order of $\mathcal{O}(t_{\max}N^3MWD)$ per step. Since the number of steps is given by $\tau_{\max} = \frac{K}{W} + D - 1$, the overall computational complexity required to estimate all MK TX symbols, *i.e.*, $\check{x}_m[k], \forall(k,m)$, is in order of $\mathcal{O}((t_{\max}D)\cdot N^3MK)$. This complexity is one exponent larger in the RX dimension N than that of the SotA Bayesian DD algorithm represented by expectation propagation (EP) [53] and may seem very large since it is proportional to the cube of N of the effective beam-domain channel $\boldsymbol{H}[k]$. Note, however, that most of the spatial DoF are used to form *pencil* beams, and therefore the effective beam-domain channel matrix size $N \times M$ is much smaller than the physical array size $N_{\mathrm{RX}} \times N_{\mathrm{TX}}$. If N is set larger than the number of major singular values of a channel matrix, *i.e.*, larger than the number of effective clusters L, substantially all available spatial DoF can be used without energy loss, with little loss of accuracy in subsequent processing. Since in an actual mmWave wireless communication channel L rarely exceeds 4 [1], N can be set sufficiently small while maintaining detection performance.

IV. PERFORMANCE ASSESSMENT

Computer simulations were conducted to validate the performance of the proposed receiver under various system conditions. The basic system format was set to follow [45], in which the carrier frequency was set to $f_c = 60$ [GHz], the DFT size was set to $N_{\mathrm{DFT}} = 512$, the guard interval was assumed to be a quarter of each OFDM symbol length, and the OFDM sampling rate was set to $f_s = 2.64$ [GHz]. The mmWave wireless communication channels considered consist of $L = 4$ clusters with $C_\ell = 15$ rays-per-cluster and varied the relative speed between the transmitter and receiver within the range of $[15, 60]$ [kmph], which is commonly used in low-speed urban scenarios of vehicle-to-everything (V2X) communication systems [15], [16].[6] Table III illustrates the values of each parameter at different relative speeds v. The evaluation at different relative speeds corresponds to verifying the detection performance of the proposed method in problem settings of different r, *i.e.*, different difficulty levels.

Simulated results are offered both in terms of BER and NMSE, averaged over 10000 different time-varying channel realizations; the number of TX/RX antennas, that of RX/TX dimensions of $\boldsymbol{H}[k]$, and that of OFDM symbols assumed to be $N_{\mathrm{TX}} = 4$, $N_{\mathrm{RX}} = 16$, $M = 2$, $N = 8$, and $K = 128$, respectively, unless otherwise specified. The damping factor was set to $\alpha = 0.5$.

A. Effects of Parameters on Performance and Convergence

Before proceeding to the presentation of performance comparison, we analyze how the internal parameters embedded in the proposed algorithm affect its performance and convergence. In the following, the parameters are set to $(W, D, G) = (8, 3, 6)$, $t_{\max} = 8$, and then each parameter is varied one by one to see how the performance changes with $v = 60$ kmph.

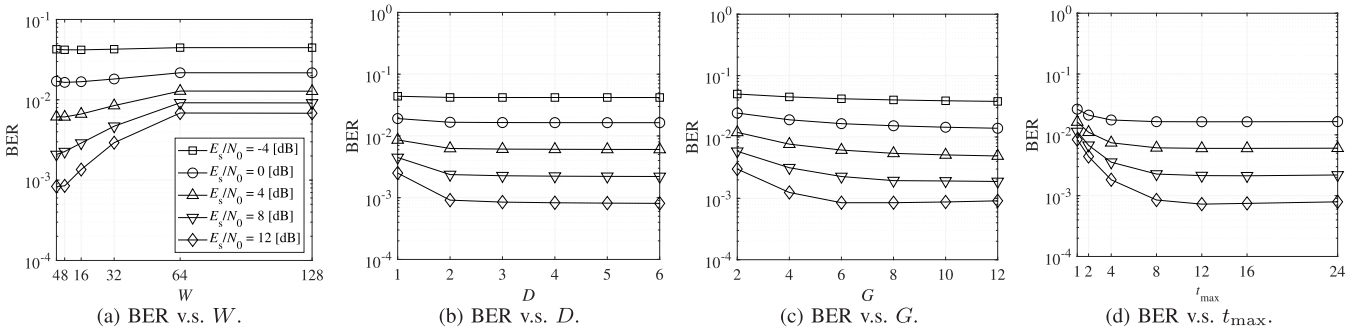
Fig. 2 shows the BER performance of the proposed JCTDD algorithm as a function of each parameter at different E_s/N_0. W is inversely proportional to the number of steps in sequential processing $\tau_{\max}$; hence, the smaller W is, the more the sequential processing strategy slides the *estimation window* little by little, which improves CT performance, but at the same time increases processing delay. Fig. 2a shows that the performance degrades as W increases, and by setting $W = 8$, sufficient detection performance can be achieved with as few steps as possible.

Next, D is the parameter that determines the number of variables estimated simultaneously for each step; hence, the complexity is proportional to D. Fig. 2b shows that the performance improvement is almost saturated at $D = 2$; hence, it is not necessary to set the value greater than 3. This fact indicates that the approximation in (19a) holds true with high accuracy as a result of the sequential process operating properly based on the Bayesian inference framework.

Next, G is the number of beliefs to be combined in VNs for CE. The larger G, the greater the number of resources used for belief combining in the discrete time dimension, which

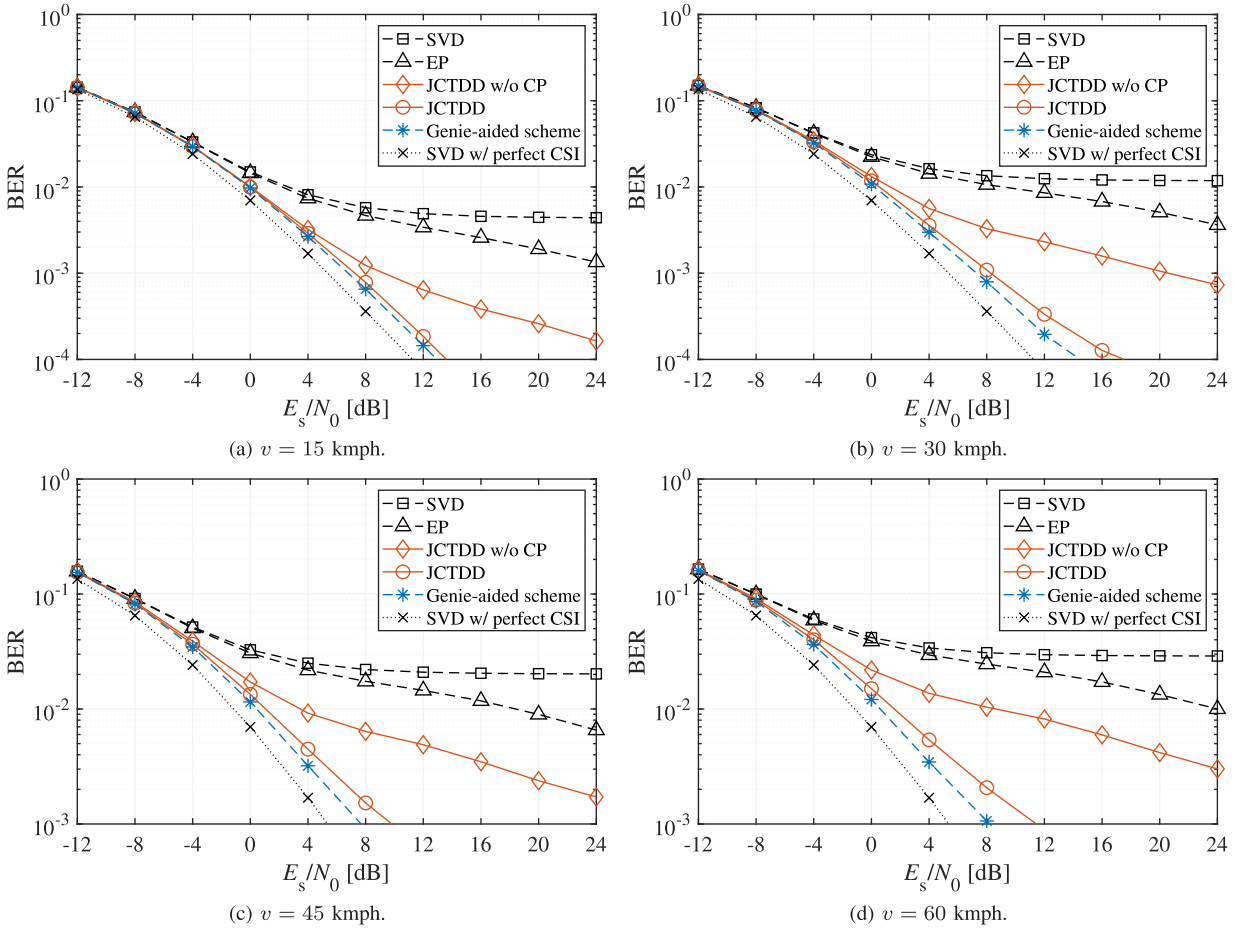
[6] In this simulation, we focus on high-mobility mmWave MIMO communication scenarios with speeds in the order of tens of kilometers per hour, which are expected in vehicle-to-vehicle (V2V) communications, vehicle-to-infrastructure (V2I) communications, as well as communications with vulnerable road users (VRUs) [54].

Note, however, that the discussion and the proposed method in this article are not restricted to any particular V2X scenario. The purpose of this simulation is to verify to what level of time variability CT can be performed by evaluating the proposed algorithm for different values of r.



(a) BER v.s. W. (b) BER v.s. D. (c) BER v.s. G. (d) BER v.s. $t_{\max}$.

Fig. 2. BER performance of time-varying mmWave MIMO systems with respect to internal parameters (W, D, G) and $t_{\max}$.



(a) $v = 15$ kmph. (b) $v = 30$ kmph.

(c) $v = 45$ kmph. (d) $v = 60$ kmph.

Fig. 3. BER performance of time-varying mmWave MIMO systems with respect to E_s/N_0.

increases the diversity gain, but at the same time, the adverse effect of errors due to channel aging increases. As expected, Fig. 2c shows that the performance improvement saturates around $G = 6$ to 8, with a slight performance degradation at higher settings, especially in the high E_s/N_0 region.

Finally, it can be seen in Fig. 2d that the performance improvement saturates around $t_{\max} = 8$ to 12. The convergence of iterative estimation is observed in the low E_s/N_0 region with fewer iterations than in the high E_s/N_0 region, indicating that detection performance can be maximized with fewer iterations, especially in the coding system.

Based on the above results, we will use $(W, D, G) = (8, 3, 6)$, $t_{\max} = 8$ in the following evaluation. The results below show that these parameters are not sensitive to various system setups, but at the choice of the system designer, further fine optimization is also possible.

B. BER Performance

Our first set of results is given in Fig. 3, where the BER performances as a function of the E_s/N_0 and for different relative velocities (*i.e.*, $v \in \{15, 30, 45, 60\}$) of the following mmWave MIMO systems are compared:

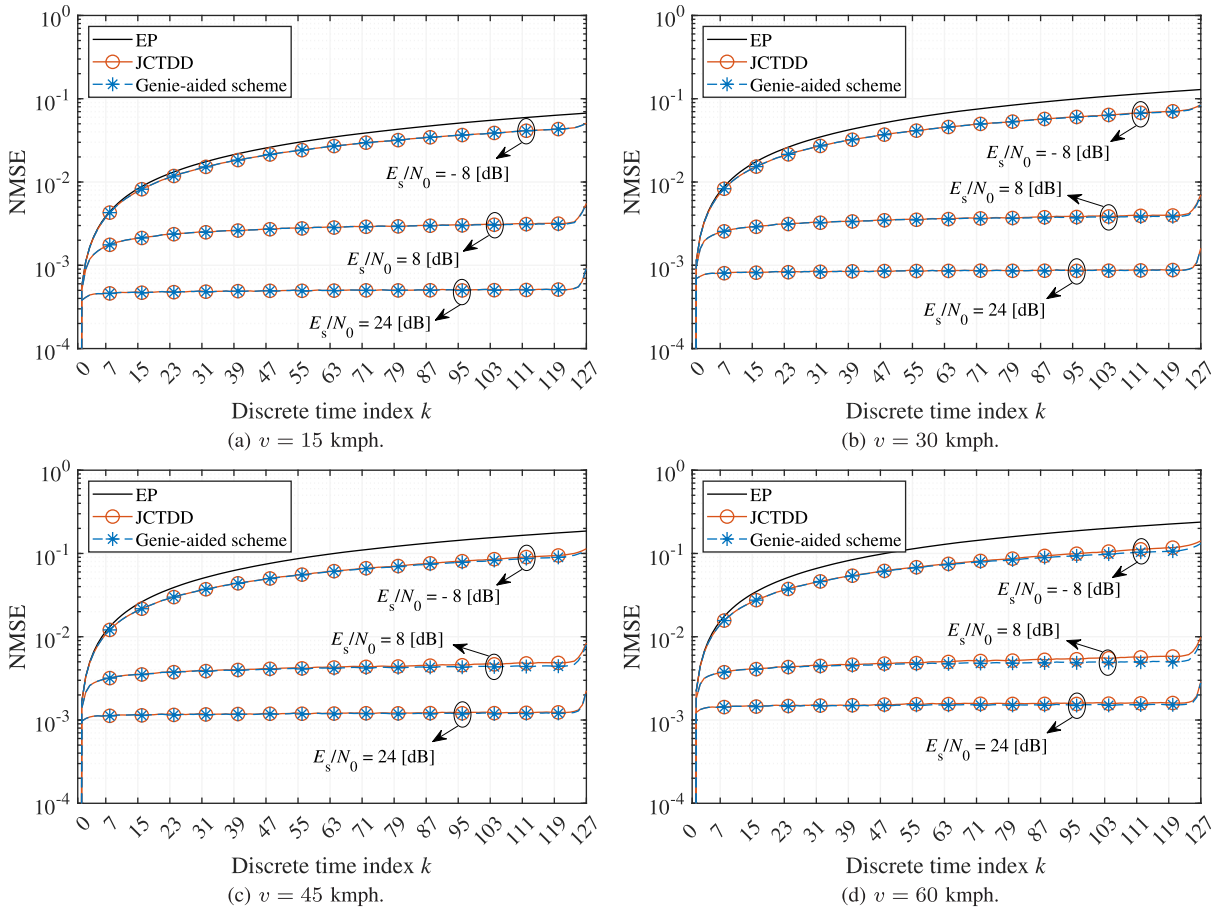


Fig. 4. NMSE performance of time-varying mmWave MIMO systems with respect to discrete time index k.

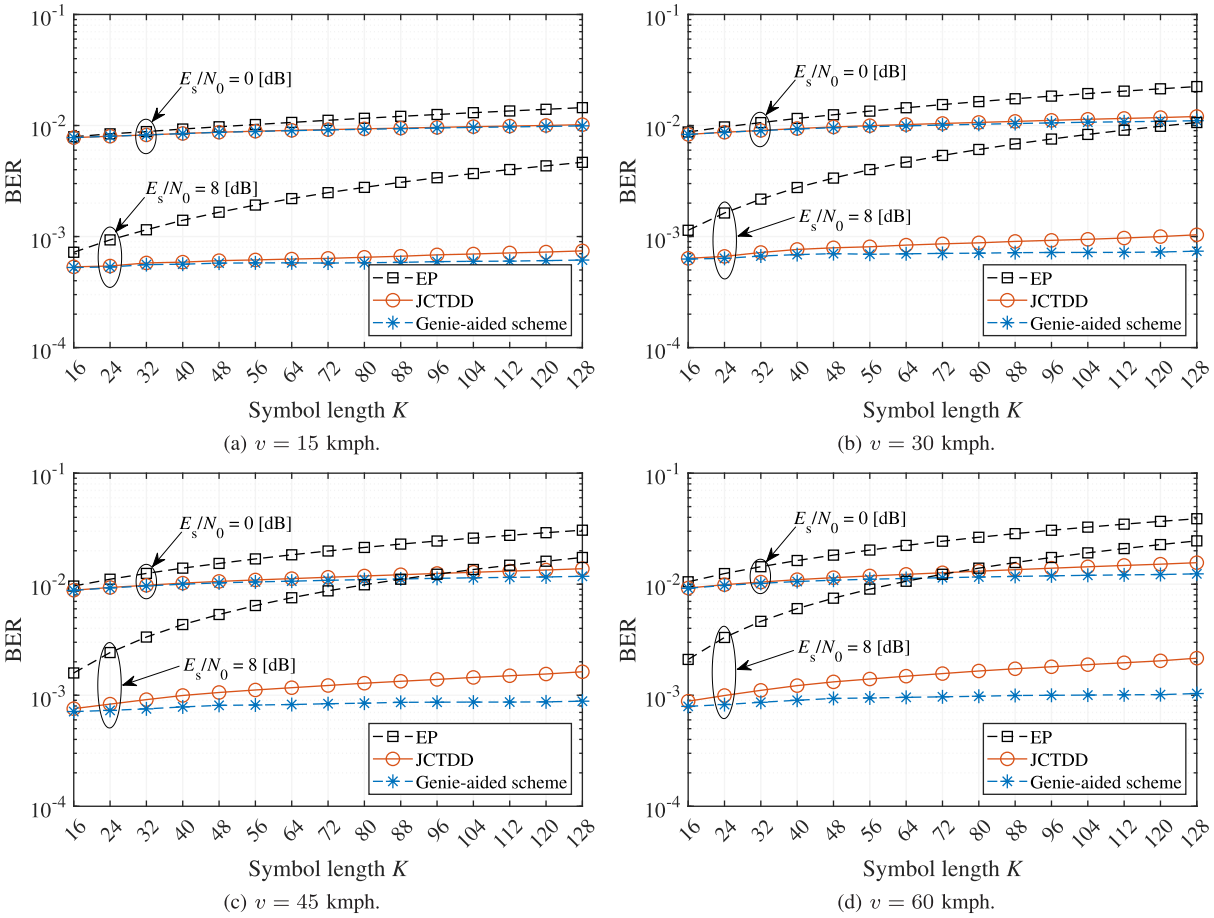
- *SVD:* Baseline TX/RX beamforming design via SVD of $\acute{\boldsymbol{H}}[0] = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathsf{H}}$, *i.e.*, $\boldsymbol{F}_{\mathrm{TX}} = [\boldsymbol{V}]_{:,1:M}$ and $\boldsymbol{F}_{\mathrm{RX}} = [\boldsymbol{U}]_{:,1:M}$, as described in Subsection II-C.
- *EP:* SotA Bayesian receiver adopting the EP-based DD algorithm [53], providing a reference performance to verify the gain achieved by the channel aging-aware JCTDD algorithm since it can achieve Bayesian optimal DD utilizing exactly the same prior knowledge as the proposed method. The number of iterations is set to 16.
- *JCTDD w/o CP:* Proposed JCTDD receiver with $(W, D, G) = (K, 1, 6)$, where the JCDE phase is applied simultaneously to K OFDM symbols without sequential processing via the CP phase. Provides a reference performance to verify the gain achieved by the CP-aided window-sliding strategy for tracking the time variability of channels.
- *JCTDD:* Proposed JCTDD receiver with $(W, D, G) = (8, 3, 6)$ presented in Algorithms 1-3.
- *Genie-aided scheme:* Idealized scheme in which perfect knowledge of $\boldsymbol{x}[k], \forall k$, is provided as prior information at the first iteration of the JCDE phase. Provides an absolute lower bound in terms of the NMSE performance of CE.
- *SVD w/ perfect CSI:* Baseline TX/RX beamforming design via SVD of $\acute{\boldsymbol{H}}[k] = \boldsymbol{U}_k\boldsymbol{\Sigma}_k\boldsymbol{V}_k^{\mathsf{H}}$, $\forall k$, *i.e.*, $\boldsymbol{F}_{\mathrm{TX}}[k] = [\boldsymbol{V}_k]_{:,1:M}$ and $\boldsymbol{F}_{\mathrm{RX}}[k] = [\boldsymbol{U}_k]_{:,1:M}$. Provides a reference lower bound, which can only be achieved in the absence of channel aging, *i.e.*, $v = 0$ kmph.

As can be seen from Fig. 3, the SotA methods, *i.e.*, SVD and EP, suffer from high error floors, which stem from the poor CE accuracy, caused by the channel aging. Even in the EP detector, which exploits the long-term channel statistics to suppress the channel aging interference, serious performance degradation is inevitable as v increases. In contrast, "JCTDD w/o CP" can improve the CE accuracy by taking advantage of the estimated data symbols as effective soft pilots, thus significantly outperforming the SotA. However, the slope of the BER curve remains gradual as in the case of EP and deviates from that of the ideal Genie-aided reference.

In comparison to the latter method, the proposed "JCTDD" method exhibits an error-floor free water-fall BER curve, approaching that of the ideal Genie-aided reference, without any additional pilots, *i.e.*, without any change in the transmission frame format. Sequential processing via the CP phase enables highly accurate CT by generating the reliable prior estimates input to the JCDE phase, reducing the degradation from the Genie-aided reference with BER $= 10^{-3}$ to about 0.5 dB, 1.0 dB, 2.0 dB, and 3.5 dB in Figs. 3a–3d, respectively.

C. MSE Performance

Next, we evaluate the CE performance of the proposed method, in terms of the NMSE of the channel estimates at



(a) $v = 15$ kmph. (b) $v = 30$ kmph. (c) $v = 45$ kmph. (d) $v = 60$ kmph.

Fig. 5. BER performance of time-varying mmWave MIMO systems with respect to symbol length K.

each discrete time $k \in \mathcal{K}$, where the $E_{\rm s}/N_0$ is fixed at -8 dB, 8 dB, and 24 dB, respectively. In view of Fig. 3, results for "SVD" and "JCTDD w/o CP" are omitted as they have shown to perform poorly, while the Genie-aided scheme has already provided a lower-bounding reference. In "EP," the channel estimate at the k-th discrete time is given by $\check{\boldsymbol{H}}[k] = r^k \boldsymbol{H}[0]$ according to the transition model in (12), which provides a lower bound without the channel tracking (CT) mechanism. With that clarified, the NMSE performances of the remaining three algorithms with the same settings as in Fig. 3 are compared in Fig. 4.

As expected, it is found that the NMSE performances of "EP" deteriorate significantly with increasing k, *i.e.*, over time, due to channel aging. In contrast, the proposed method asymptotically approaches the Genie-aided scheme, and only slight deviations from the reference are observed even under the high-mobility communication scenarios in Figs. 4c and 4d. Furthermore, in higher $E_{\rm s}/N_0$ regime, *i.e.*, at 8 dB and 24 dB, NMSEs are almost constant with respect to k, confirming that highly accurate CT can be achieved[7].

[7]The increase in NMSE around $k = 127$ is due to the fact that the number of beliefs that contribute to the diversity gain in the belief combining of (39) decreases owing to the unavailability of observations at future times $k > 127$, and this inconvenience can be avoided by properly terminating the process using known control signals in practical systems.

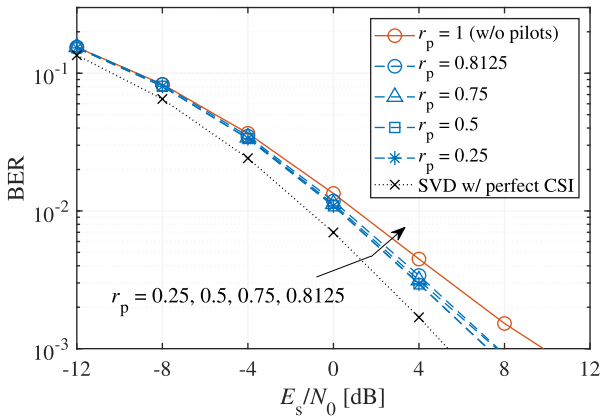
D. Insight on Error Propagation in Channel Tracking

To gain insight into the harmful effect of error propagation in the proposed method, the BER performance as a function of the symbol length K with $E_{\rm s}/N_0 = 0$ dB and 8 dB are compared in Fig. 5, for the same mobility scenarios of Fig. 4.

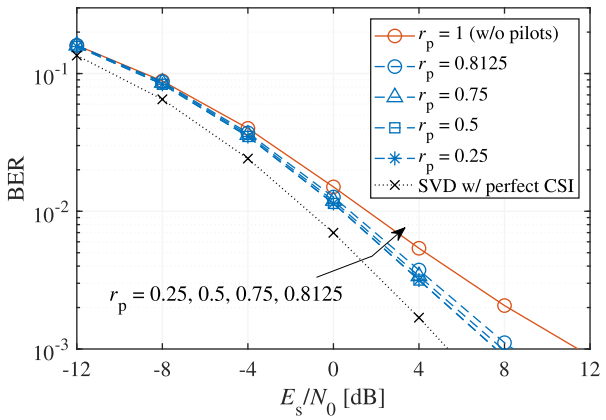
As expected from the poor NMSE performance shown in Fig. 4, the performance of "EP" deteriorates rapidly as K increases, even with the DD algorithm to account for the channel aging interference. In contrast, the proposed method achieves significant performance improvement due to its highly accurate CT mechanism. However, particularly in scenarios with large values of v as shown in Figs. 5c and 5d, it is found that the deviation from the performance of the Genie-aided scheme increases with increasing K, albeit quite slowly. This performance degradation can be attributed to the error propagation in sequential processing, and some sort of error correction mechanism must be introduced separately to compensate for it. With that in mind, the following subsections study the effect of inserting additional pilot symbols and using error-correcting codes on system performance, respectively, providing insight into the optimal system configuration.

E. Additional Pilot Insertion

In this subsection, we study the effect of inserting additional pilot symbols. We here define the pilot rate $r_{\rm p}$ as the proportion

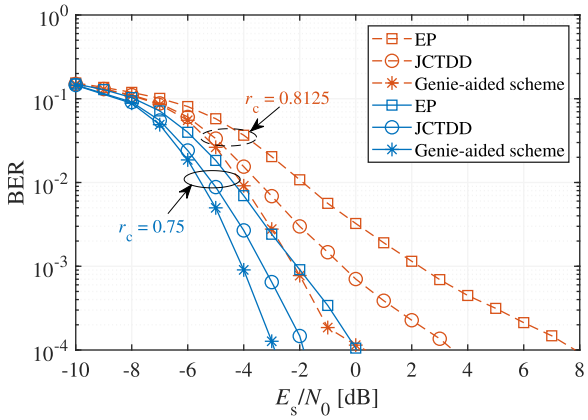


(a) $v = 45$ kmph.

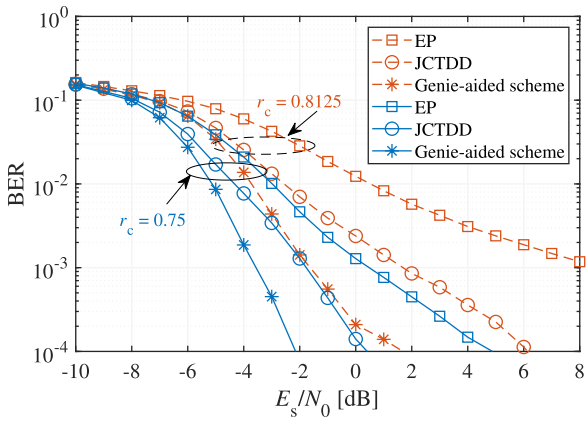


(b) $v = 60$ kmph.

Fig. 6. BER performance of the proposed method with respect to E_s/N_0.



(a) $v = 45$ kmph.



(b) $v = 60$ kmph.

Fig. 7. BER performance of coded mmWave MIMO systems with respect to E_s/N_0, where the dashed and solid lines indicate performance for $r_c = 0.8125$ and 0.75, respectively.

of data symbols to the symbol length K. That is, when the number of embedded pilot symbols is K_p, $r_p \triangleq 1 - K_p/K$. The K_p pilot symbols given by the rows of a $K_p \times K_p$ DFT matrix are assumed to be placed as evenly spaced as possible to maximize the error correction capability during the iterative estimation process.

Fig. 6 shows the BER performances of the proposed method for different pilot rates, *i,e,*, $r_p \in \{1, 0.8125, 0.75, 0.5, 0.25\}$, with the same settings as in Figs. 3c and 3d. From both figures, it can be confirmed that periodically embedded pilot symbols function as a kind of error correction mechanism and suppress error propagation, resulting in improved performance. It is worth noting here that while a small number of pilot symbols provides reasonable performance improvement, excessive piloting (*i.e.*, a lower pilot rate) does not result in corresponding improvement. In fact, the deviation between the performances of $r_p = 0.8125$ and $r_p = 0.25$ is less than 1.0 dB at BER $= 10^{-3}$, and the improvement is almost saturated at $r_p = 0.75$. This performance is the lower bound that can be achieved by adding pilot symbols, and it can be seen that it is roughly consistent with that of the idealized Genie-aided reference in Fig. 3. These results indicate that in the proposed JCTDD algorithm, the performance gains from adding pilot symbols are limited and that the number of additional pilot symbols, if any, should be small. Furthermore, the existing pilot-based counterparts [20], [21], [22] are unlikely to achieve performance equal to or better than the saturation performance ($r_p = 0.25$) since they perform CT/CE based only on reference signaling, suggesting that the proposed algorithm can at least achieve asymptotic performance approaching that of the existing methods, even without an additional piloting ($r_p = 1$).

F. BER Performance in Coded Systems

Finally, in this subsection we compare the BER performances of coded mmWave MIMO systems adopting the low-density parity-check (LDPC) code with rates $r_c \in \{0.8125, 0.75\}$, in alignment with fifth-generation (5G) new radio (NR) specification [55], where the sum-product algorithm (SPA) is employed as the soft-decision decoder. Remarkably, it is found in Fig. 7 that the proposed method exhibits error-floor-free water-fall BER curves in the E_s/N_0 region observed, despite the relatively high coding rates utilized, namely, $r_c = 0.8125$ and $r_c = 0.75$.

However, due to the reduction of the Gaussianity of log-likelihood ratios (LLRs) used as input in the channel decoder, which is caused by interference due to channel aging, a deviation from the ideal Genie-aided reference is also observed, especially at high coding rates, although the gap becomes smaller as r_c decreases. Large gains compared to the EP-based

DD method are obtained in all figures, confirming that the proposed method can appropriately benefit from the coding gains from the error correction process.

Comparing the BER performances of the proposed method at a given transmission rate, as shown in Figs. 6 and 7 for $r_{\mathrm{p}} = r_{\mathrm{c}} = 0.75$ and $r_{\mathrm{p}} = r_{\mathrm{c}} = 0.8125$, it can be seen that rather than adding pilot symbols to aid CT, such redundancy is better employed to boost error correction capability via channel coding, which leads to greater performance improvement.

V. Conclusion

We proposed a novel BBI-based JCTDD receiver for uplink mmWave MIMO systems with time-varying channels. Taking advantage of the sequential two-step tracking scheme found in the typical KF as the basic framework, new belief generation and belief combining processes were designed by incorporating the AR model, which describes the time variability of wireless channels, into the JCDE algorithm as the state transition model. Under consistent assumptions and approximation accuracy, a new overall belief propagation regime was designed according to the BiGaBP framework, and only the DD mechanism was designed based on the PDA-like approach to account for the spatial correlation among fading coefficients in mmWave wireless channels. Finally, computer simulations were conducted to demonstrate the validity of our proposed method in terms of BER and NMSE performances in various system setups. The results show that the proposed method can continue to track channels with high accuracy without any additional pilots. Furthermore, comparative study with the SotA alternatives and the case studies covering various scenarios provided insight into the optimal system design.

Appendix
Second-Order Statistics of Errors Due to Channel-Aging Phenomenon

Based on (12), the covariance matrix of errors due to channel aging can be in general obtained from the conditional expectation of $\boldsymbol{H}[k]\boldsymbol{H}[k]^{\mathsf{H}}$, given $\boldsymbol{H}[k-k']$ as

$$\begin{aligned}
&\boldsymbol{\Omega}_{k'} \\
&\triangleq \mathbb{E}_{\mathbf{H}[k]|\mathbf{H}[k-k']}\left[\boldsymbol{H}[k]\boldsymbol{H}[k]^{\mathsf{H}}\middle|\boldsymbol{H}[k-k']\right] \\
&= \mathbb{E}\left[\left(\boldsymbol{H}[k]-r^{k'}\boldsymbol{H}[k-k']\right)\left(\boldsymbol{H}[k]-r^{k'}\boldsymbol{H}[k-k']\right)^{\mathsf{H}}\right] \\
&= \frac{1-r^{2k'}}{L}\sum_{l=1}^{L}\sum_{c=1}^{C_\ell}\frac{\boldsymbol{A}_{\ell,c}\boldsymbol{A}_{\ell,c}^{\mathsf{H}}}{C_\ell}. \qquad (44a)
\end{aligned}$$

Similarly, the conditional covariance matrix of channel vector corresponding to each user equipment (UE) can be calculated as

$$\begin{aligned}
&\boldsymbol{\Omega}_{k',m} \\
&\triangleq \mathbb{E}_{\mathbf{h}_m[k]|\mathbf{h}_m[k-k']}\left[\boldsymbol{h}_m[k]\boldsymbol{h}_m[k]^{\mathsf{H}}\middle|\boldsymbol{h}_m[k-k']\right] \\
&= \mathbb{E}\left[\left(\boldsymbol{h}_m[k]-r^{k'}\boldsymbol{h}_m[k-k']\right)\left(\boldsymbol{h}_m[k]-r^{k'}\boldsymbol{h}_m[k-k']\right)^{\mathsf{H}}\right] \\
&= \frac{1-r^{2k'}}{L}\sum_{l=1}^{L}\sum_{c=1}^{C_\ell}\frac{[\boldsymbol{A}_{\ell,c}]_{:,m}[\boldsymbol{A}_{\ell,c}]_{:,m}^{\mathsf{H}}}{C_\ell}, \qquad (44b)
\end{aligned}$$

and the conditional element-wise MSE can be calculated as

$$\begin{aligned}
&\omega_{k',nm} \\
&\triangleq \mathbb{E}_{\mathsf{h}_{nm}[k]|\mathsf{h}_{nm}[k-k']}\left[|h_{nm}[k]|^2\middle|h_{nm}[k-k']\right] \\
&= \mathbb{E}\left[\left(h_{nm}[k]-r^{k'}h_{nm}[k-k']\right)\left(h_{nm}[k]-r^{k'}h_{nm}[k-k']\right)^{*}\right] \\
&= \frac{1-r^{2k'}}{L}\sum_{l=1}^{L}\sum_{c=1}^{C_\ell}\frac{|[\boldsymbol{A}_{\ell,c}]_{n,m}|^2}{C_\ell}. \qquad (44c)
\end{aligned}$$

References

[1] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.

[2] C.-X. Wang, J. Bian, J. Sun, W. Zhang, and M. Zhang, "A survey of 5G channel measurements and models," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3142–3168, 4th Quart., 2018.

[3] J. Huang, C.-X. Wang, H. Chang, J. Sun, and X. Q. Gao, "Multi-frequency multi-scenario millimeter wave MIMO channel measurements and modeling for B5G wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 9, pp. 2010–2025, Sep. 2020.

[4] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing—The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[5] 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.

[6] J. Lee, G.-T. 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.

[7] Z. Xiao, P. Xia, and X.-G. Xia, "Codebook design for millimeter-wave channel estimation with hybrid precoding structure," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 141–153, Jan. 2017.

[8] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive MIMO: The next wireless revolution?" *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 56–62, Sep. 2014.

[9] I. K. Jain, R. Kumar, and S. S. Panwar, "The impact of mobile blockers on millimeter wave cellular systems," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 4, pp. 854–868, Apr. 2019.

[10] Y. Liu, C.-X. Wang, J. Huang, J. Sun, and W. Zhang, "Novel 3-D nonstationary mmWave massive MIMO channel models for 5G high-speed train wireless communications," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 2077–2086, Mar. 2019.

[11] T. Nishio et al., "Proactive received power prediction using machine learning and depth images for mmWave networks," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 11, pp. 2413–2427, Nov. 2019.

[12] H. Iimori, G. T. F. de Abreu, O. Taghizadeh, R.-A. Stoica, T. Hara, and K. Ishibashi, "Stochastic learning robust beamforming for millimeter-wave systems with path blockage," *IEEE Wireless Commun. Lett.*, vol. 9, no. 9, pp. 1557–1561, Sep. 2020.

[13] K. Rikkinen, P. Kyösti, M. E. Leinonen, M. Berg, and A. Pärssinen, "THz radio communication: Link budget analysis toward 6G," *IEEE Commun. Mag.*, vol. 58, no. 11, pp. 22–27, Nov. 2020.

[14] J. Choi, V. Va, N. Gonzalez-Prelcic, R. Daniels, C. R. Bhat, and R. W. Heath Jr., "Millimeter-wave vehicular communication to support massive automotive sensing," *IEEE Commun. Mag.*, vol. 54, no. 12, pp. 160–167, Dec. 2016.

[15] N. Ishikawa et al., "Differential-detection aided large-scale generalized spatial modulation is capable of operating in high-mobility millimeter-wave channels," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 6, pp. 1360–1374, Oct. 2019.

[16] H. Iimori, G. T. F. de Abreu, and O. Gonsa, "Mitigating channel aging and phase noise in millimeter wave MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 70, no. 7, pp. 7237–7242, Jul. 2021.

[17] N. González-Prelcic, H. Xie, J. Palacios, and T. Shimizu, "Wide-band channel tracking and hybrid precoding for mmWave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 4, pp. 2161–2174, Apr. 2021.

[18] J. Seo, Y. Sung, G. Lee, and D. Kim, "Training beam sequence design for millimeter-wave MIMO systems: A POMDP framework," *IEEE Trans. Signal Process.*, vol. 64, no. 5, pp. 1228–1242, Mar. 2016.

[19] S. H. Lim, S. Kim, B. Shim, and J. W. Choi, "Efficient beam training and sparse channel estimation for millimeter wave communications under mobility," *IEEE Trans. Commun.*, vol. 68, no. 10, pp. 6583–6596, Oct. 2020.

[20] Q. Qin, L. Gui, P. Cheng, and B. Gong, "Time-varying channel estimation for millimeter wave multiuser MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9435–9448, Oct. 2018.

[21] L. Cheng, G. Yue, X. Xiong, Y. Liang, and S. Li, "Tensor decomposition-aided time-varying channel estimation for millimeter wave MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 4, pp. 1216–1219, Aug. 2019.

[22] J. Wang, W. Zhang, Y. Chen, Z. Liu, J. Sun, and C.-X. Wang, "Time-varying channel estimation scheme for uplink MU-MIMO in 6G systems," *IEEE Trans. Veh. Technol.*, vol. 71, no. 11, pp. 11820–11831, Nov. 2022.

[23] J. Ma, S. Zhang, H. Li, F. Gao, and S. Jin, "Sparse Bayesian learning for the time-varying massive MIMO channels: Acquisition and tracking," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1925–1938, Mar. 2019.

[24] J. H. Yoo, J. Bae, S. H. Lim, S. Kim, J. W. Choi, and B. Shim, "Sampling-based tracking of time-varying channels for millimeter wave-band communications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.

[25] S. Kashyap, C. Mollen, E. Bjornson, and E. G. Larsson, "Performance analysis of (TDD) massive MIMO with Kalman channel prediction," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 3554–3558.

[26] H. Kim, S. Kim, H. Lee, C. Jang, Y. Choi, and J. Choi, "Massive MIMO channel prediction: Kalman filtering vs. machine learning," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 518–528, Jan. 2021.

[27] J. Yuan, H. Q. Ngo, and M. Matthaiou, "Machine learning-based channel prediction in massive MIMO with channel aging," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 2960–2973, May 2020.

[28] G. Fodor, S. Fodor, and M. Telek, "MU-MIMO receiver design and performance analysis in time-varying Rayleigh fading," *IEEE Trans. Commun.*, vol. 70, no. 2, pp. 1214–1228, Feb. 2022.

[29] J. T. Parker, P. Schniter, and V. Cevher, "Bilinear generalized approximate message passing—Part I: Derivation," *IEEE Trans. Signal Process.*, vol. 62, no. 22, pp. 5839–5853, Nov. 2014.

[30] C.-K. Wen, C.-J. Wang, S. Jin, K.-K. Wong, and P. Ting, "Bayes-optimal joint channel-and-data estimation for massive MIMO with low-precision ADCs," *IEEE Trans. Signal Process.*, vol. 64, no. 10, pp. 2541–2556, May 2016.

[31] L. Chen and X. Yuan, "Blind multiuser detection in massive MIMO channels with clustered sparsity," *IEEE Wireless Commun. Lett.*, vol. 8, no. 4, pp. 1052–1055, Aug. 2019.

[32] S. Sarkar, A. K. Fletcher, S. Rangan, and P. Schniter, "Bilinear recovery using adaptive vector-AMP," *IEEE Trans. Signal Process.*, vol. 67, no. 13, pp. 3383–3396, Jul. 2019.

[33] K. Ito, T. Takahashi, S. Ibi, and S. Sampei, "Bilinear Gaussian belief propagation for massive MIMO detection with non-orthogonal pilots," *IEEE Trans. Commun.*, vol. 72, no. 2, pp. 1045–1061, Feb. 2024.

[34] H. Iimori, T. Takahashi, K. Ishibashi, G. T. F. de Abreu, and W. Yu, "Grant-free access via bilinear inference for cell-free MIMO with low-coherence pilots," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7694–7710, Nov. 2021.

[35] H. Iimori, T. Takahashi, K. Ishibashi, G. T. F. de Abreu, D. G. González, and O. Gonsa, "Joint activity and channel estimation for extra-large MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 7253–7270, Sep. 2022.

[36] T. Takahashi, H. Iimori, K. Ando, K. Ishibashi, S. Ibi, and G. T. F. de Abreu, "Bayesian receiver design via bilinear inference for cell-free massive MIMO with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 22, no. 7, pp. 4756–4772, Jul. 2023.

[37] Z. Xiao, X.-G. Xia, D. Jin, and N. Ge, "Iterative eigenvalue decomposition and multipath-grouping Tx/Rx joint beamformings for millimeter-wave communications," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1595–1607, Mar. 2015.

[38] R. W. Heath Jr., N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[39] 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.

[40] R.-A. Stoica, H. Iimori, G. T. Freitas de Abreu, and K. Ishibashi, "Frame theory and fractional programming for sparse recovery-based mmWave channel estimation," *IEEE Access*, vol. 7, pp. 150757–150774, 2019.

[41] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath Jr., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[42] T. S. Rappaport, *Wireless Communications—Principles and Practice.* Upper Saddle River, NJ, USA: Prentice-Hall, 1996.

[43] V. Va, J. Choi, and R. W. Heath Jr., "The impact of beamwidth on temporal channel variation in vehicular channels and its implications," *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, pp. 5014–5029, Jun. 2017.

[44] Y. Kang, H. Seo, and W. Choi, "When to realign the receive beam in high mobility V2X communications?" *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 13180–13195, Nov. 2020.

[45] C. Cordeiro, D. Akhmetov, and M. Park, "IEEE 802.11ad: Introduction and performance evaluation of the first multi-Gbps WiFi technology," in *Proc. ACM Int. Workshop mmWave Commun. Circuits Netw.*, New York, NY, USA, Sep. 2010, pp. 3–8.

[46] A. A. Zaidi et al., "Waveform and numerology to support 5G services and requirements," *IEEE Commun. Mag.*, vol. 54, no. 11, pp. 90–98, Nov. 2016.

[47] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.

[48] S. Yang, T. Lv, R. G. Maunder, and L. Hanzo, "From nominal to true a posteriori probabilities: An exact Bayesian theorem based probabilistic data association approach for iterative MIMO detection and decoding," *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 2782–2793, Jul. 2013.

[49] A. Chockalingam and B. S. Rajan, *Large MIMO Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2014.

[50] T. Takahashi, A. Tölli, S. Ibi, and S. Sampei, "Low-complexity large MIMO detection via layered belief propagation in beam domain," *IEEE Trans. Wireless Commun.*, vol. 21, no. 1, pp. 234–249, Jan. 2022.

[51] L. Hanzo, T. H. Liew, B. L. Yeap, R. Y. S. Tee, and S. X. Ng, *Turbo Coding, Turbo Equalisation and Space-Time Coding: EXIT-Chart-Aided Near-Capacity Designs for Wireless Channels*. Hoboken, NJ, USA: Wiley, Mar. 2011.

[52] T. Takahashi, S. Ibi, and S. Sampei, "Design of adaptively scaled belief in multi-dimensional signal detection for higher-order modulation," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1986–2001, Mar. 2019.

[53] K. Takeuchi, "Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements," *IEEE Trans. Inf. Theory*, vol. 66, no. 1, pp. 368–386, Jan. 2020.

[54] M. Noor-A-Rahim et al., "6G for vehicle-to-everything (V2X) communications: Enabling technologies, challenges, and opportunities," *Proc. IEEE*, vol. 110, no. 6, pp. 712–734, Jun. 2022.

[55] 3GPP, *5G NR Multiplexing and Channel Coding*, document (TS) 38.212, 2018.

Takumi Takahashi (Member, IEEE) received the B.E., M.E., and Ph.D. degrees in communication engineering from Osaka University, Osaka, Japan, in 2016, 2017, and 2019, respectively. From 2018 to 2019, he was a Visiting Researcher with the Centre for Wireless Communications, University of Oulu, Finland. In 2019, he joined the Graduate School of Engineering, Osaka University, as an Assistant Professor. His current research interests include Bayesian inference, belief propagation, signal processing, and wireless communications.

Hiroki Iimori (Member, IEEE) received the B.Eng. and M.Eng. (Hons.) degrees in electrical and electronic engineering from Ritsumeikan University, Kyoto, Japan, in 2017 and 2019, respectively, and the Ph.D. degree (Hons.) from Jacobs University Bremen, Bremen, Germany, in 2022. In 2020, he was a Visiting Scholar with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada. In 2021, he was a Research Intern with the Ericsson Radio S&R Research Laboratory, Yokohama, Japan, where he is currently an experienced Researcher. His research interests include optimization theory, wireless communications, and signal processing. He received the YKK Doctoral Fellowship from the Yoshida Scholarship Foundation, the IEICE Young Researcher of the Year Award from the IEICE Smart Radio Committee in 2020, and among others.

Koji Ishibashi (Senior Member, IEEE) received the B.E. and M.E. degrees in engineering from The University of Electro-Communications, Tokyo, Japan, in 2002 and 2004, respectively, and the Ph.D. degree in engineering from Yokohama National University, Yokohama, Japan, in 2007. From 2007 to 2012, he was an Assistant Professor with the Department of Electrical and Electronic Engineering, Shizuoka University, Hamamatsu, Japan. Since April 2012, he has been with the Advanced Wireless and Communication Research Center (AWCC), The University of Electro-Communications, where he is currently a Professor. From 2010 to 2012, he was a Visiting Scholar with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA. His current research interests include robust beamforming, grant-free access, energy harvesting, compressed sensing, coding, cell-free networks, and MIMO technologies. He is a Senior Member of IEICE. He was a recipient of the Takayanagi Research Encouragement Award in 2009 and the KDDI Foundation Award in 2023. He was certified as an Exemplary Reviewer of IEEE COMMUNICATIONS WIRELESS LETTERS in 2015 and awarded by the Telecommunication Technology Committee (TTC) for his devotion to standardization activities in 2020. He was an Associate Editor of *IEICE Transactions on Communications* and IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS and a Guest Editor of IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY.

Shinsuke Ibi (Member, IEEE) received the B.E. degree in advanced engineering from Suzuka College of Technology, Japan, in 2002, and the M.E. and Ph.D. degrees in communication engineering from Osaka University, Japan, in 2004 and 2006, respectively. From 2005 to 2006, he was a Visiting Researcher with the Centre for Wireless Communications, University of Oulu, Finland. In 2006, he joined the Graduate School of Engineering, Osaka University. From 2010 to 2011, he was a Visiting Researcher with the University of Southampton, U.K. He moved to Doshisha University in 2019, where he is currently a Professor with the Faculty of Science and Engineering. His research interests include EXIT-based coding theory, iterative detection, digital signal processing, cognitive radio, and communication theory. He received the 64th and 71st Best Paper Awards from IEICE and the 24th Telecom System Technology Award from the Telecommunication Advancement Foundation.

Giuseppe Thadeu Freitas de Abreu (Senior Member, IEEE) received the B.Eng. degree in electrical engineering and a specialization (Latu Sensu) degree in telecommunications engineering from Universidade Federal da Bahia (UFBa), Salvador, Bahia, Brazil, in 1996 and 1997, respectively, and the M.Eng. and D.Eng. degrees in physics, electrical and computer engineering from Yokohama National University, Japan, in March 2001 and March 2004, respectively. He was a Post-Doctoral Fellow and later an Adjunct Professor (Docent) of statistical signal processing and communications theory with the Department of Electrical and Information Engineering, University of Oulu, Finland, from 2004 to 2006 and from 2006 to 2011, respectively. Since 2011, he has been a Professor of electrical engineering with Constructor University (formerly Jacobs University Bremen), Germany. From April 2015 to August 2018, he was a Full Professor with the Department of Computer and Electrical Engineering, Ritsumeikan University, Japan. His research interests include communications and signal processing, including communications theory, estimation theory, statistical modeling, wireless localization, cognitive radio, wireless security, MIMO systems, ultrawideband and millimeter wave communications, full-duplex and cognitive radio, compressive sensing, energy harvesting networks, random networks, connected vehicles networks, joint communications and sensing, and many others. He was a recipient of the Uenohara Award from Tokyo University in 2000 for the master's thesis work. He was a co-recipient of best paper awards at several international conferences and was awarded the JSPS, Heiwa Nakajima, and NICT Fellowships (twice) in 2010, 2013, 2015, and 2018, respectively. He served as an Associate Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2009 to 2014 and IEEE TRANSACTIONS ON COMMUNICATIONS from 2014 to 2017 and an Executive Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2018 to 2021. He is an Editor of IEEE SIGNAL PROCESSING LETTERS and IEEE COMMUNICATIONS LETTERS.