Compressed Sensing-Aided Downlink Channel Training for FDD Massive MIMO Systems

Yonghee Han, *Member, IEEE*, Jungwoo Lee, *Senior Member, IEEE*, and David J. Love, *Fellow, IEEE*

***Abstract*— There is much discussion in industry and academia about possible technical solutions to address the growth in demand for wireless broadband. Massive multiple-input multiple-output (MIMO) systems are one of the most popular solutions to addressing this broadband demand in fifth generation (5G) cellular systems. Massive MIMO systems employ tens or hundreds of antennas at the base station to enable advanced multiuser MIMO communications. To reap the massive MIMO throughput gain, coherent transmission exploiting accurate channel state information at the transmitter is required. While it is expected that many 5G systems will employ frequency division duplexing (FDD), channel sounding for FDD systems requires a large pilot overhead, which usually scales proportionally to the number of transmit antennas. To resolve this problem, a compressed sensing (CS)-aided channel estimation scheme is proposed, which exploits the observation that the channel statistics change slowly in time. By utilizing a conventional least squares approach and a CS technique simultaneously, the proposed scheme reduces the pilot overhead. Simulation results show that the proposed scheme can estimate the channel with a reduced pilot overhead even when conventional CS cannot be applied.**

***Index Terms*— Massive MIMO, frequency division duplex, channel estimation, compressed sensing.**

I. INTRODUCTION

TO MEET the demand for high throughput in next generation (e.g. 5G) cellular networks, various directions for physical layer evolution are being explored [1]. Among them, massive multiple-input multiple-output (MIMO) systems, characterized by a large number of antennas at the transmitter, are drawing significant interest for potential standardization.

Since the seminal analysis in [2], work has shown that massive MIMO systems can provide high throughput and energy efficiency improvements with simple transmission/reception techniques [2], [3]. However, such advantages are based upon the premise of accurate channel state information (CSI) at the transmitter (CSIT). Due to the large dimension of the channels, CSIT acquisition is one of the most challenging problems in massive MIMO systems. To exploit the channel reciprocity and acquire CSIT with an uplink training sequence whose length is not proportional to the number of downlink transmit antennas, most massive MIMO research has assumed two-way communication is facilitated through time division duplexing (TDD) [2]–[4]. However, the complicated calibration required for TDD reciprocity and the adoption of frequency division duplexing (FDD) in most of the contemporary cellular networks make FDD an attractive option for 5G networks.

Accordingly, CSIT acquisition in FDD massive MIMO systems is of great importance. In FDD systems, CSIT can be achieved through downlink training and feedback from the receiver. Due to the large dimension of the channel, both the channel training and feedback are challenging in massive MIMO systems. On the feedback issue, some techniques have been proposed to reduce the feedback overhead and encoding complexity by leveraging structured quantizers such as trellis-coded quantization [5]–[7] or projecting the channel into a lower dimensional subspace [8], [9]. CSIT feedback, however, hinges on accurate downlink channel estimation, which is the focus of this paper.

Conventional channel estimation schemes adopted for small-scale MIMO systems, such as least squares (LS)-based schemes, require the length of the training sequence to be proportional to the number of transmit antennas and become inefficient in massive MIMO systems. To tackle this problem, some recent work has been devoted to estimating a massive MIMO channel with a short training sequence exploiting the correlated nature of the massive MIMO channel [10], [11]. In [10], Kalman filter-based channel estimation was studied and a pilot design minimizing the mean squared error (MSE) was developed. With this approach, the channel can be estimated using a short training sequence given transmitter knowledge of the previous channel outputs received during training or channel statistics. However, acquiring the channel statistics for non-stationary massive MIMO channels requires additional feedback. While training techniques proposed in [11] can be performed without the transmitter's knowledge of channel statistics, these schemes are practically effective only for slow-fading channels with high temporal correlation since they are based on adaptive filtering.

Compressed sensing (CS) is an attractive strategy to estimate a sparse channel with a short sequence when the channel statistics are unknown. According to the CS theory, a sparse vector can be recovered from a compressive

Manuscript received June 19, 2016; revised November 8, 2016 and February 5, 2017; accepted March 27, 2017. Date of publication April 6, 2017; date of current version July 13, 2017. This research was supported in part by the Basic Science Research Programs (NRF-2015R1A2A1A15052493) through NRF funded by MSIP, the Technology Innovation Program (10051928) funded by MOTIE, the Bio-Mimetic Robot Research Center funded by DAPA (UD130070ID), and BK21-plus. The associate editor coordinating the review of this paper and approving it for publication was L. Cottatellucci. *(Corresponding author: Jungwoo Lee.)*

Y. Han and J. Lee are with the Institute of New Media and Communications, Department of Electrical and Computer Engineering, Seoul National University, Seoul 08826, South Korea (e-mail: yhan@wspl.snu.ac.kr; junglee@snu.ac.kr).

D. J. Love is with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907 USA (e-mail: djlove@purdue.edu).

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

Digital Object Identifier 10.1109/TCOMM.2017.2691700

measurement [12]–[15], and a sparse channel can be estimated efficiently leveraging the CS theory [16], [17]. Since massive MIMO channels exhibit limited scatters compared to the large number of antennas, it can be represented by a sparse vector in the virtual angular domain [18], [19]. Exploiting the sparsity, CS has been utilized for the estimation [20], [21] and feedback design [22] for massive MIMO channels. In [20], channel estimation using a low-rank approximation was proposed for TDD multiuser massive MIMO systems. In [21], a CS-based channel estimation method was proposed, which exploits the joint sparsity of the channel. However, considering that about $O(k\log(M))$ measurements are required for the recovery of a k-sparse length-M vector [12], direct application of CS gives very little training overhead reduction with the sparsity levels prevalent in non-millimeter (centimeter or larger) wave massive MIMO channels. It is worth mentioning that since millimeter wave channels tend to have only a few significant paths and directional beamforming with a large number of antennas is essentially required to mitigate the severe attenuation [23], CS-based channel estimation is more suitable for millimeter wave systems [24]–[26].

In this paper, we propose a novel channel estimation scheme for FDD massive MIMO systems, which combines LS and CS techniques. A key observation for the proposed scheme is that even though the channel statistics change with time, the rate of change is much lower than that of the channel response. Since the variation in channel statistics occurs with this appearance and disappearance of scatterers at specific locations, the channel space intuitively evolves slowly if parameterized correctly. This approach has previously been modeled as a birth-death evolution of scattering clusters in [27]. From a CS viewpoint, this means that the support of the previous channel can be a good predictor for the support of the current channel. In other words, only a small number of elements are nonzero outside the support of the previous channel. Hence, it is possible to separate the channel vectors into a dense vector (by projecting into the previous support) and a sparse vector (by projecting into the null space of the previous support). The two vectors are estimated with different strategies. The LS and CS techniques are used to estimate the dense and sparse vectors, respectively. With the separation, the proposed scheme can reduce pilot overhead and improve tracking of the channel subspace.

The rest of this paper is organized as follows. In Section II, the channel model and the downlink channel estimation problem are introduced. In Section III, the CS-aided channel estimation scheme is proposed for FDD massive MIMO systems, and some practical issues are discussed in Section IV. In Section V, numerical results are provided. Finally, conclusions are drawn in Section VI.

Notation: A vector (matrix) is written in boldface as $\mathbf{a}$ ($\mathbf{A}$). $\mathbf{a}^*$ ($\mathbf{A}^*$), $\mathbf{a}^T$ ($\mathbf{A}^T$), and $\mathbf{a}^H$ ($\mathbf{A}^H$) denote the complex conjugate, the transpose, and the conjugate transpose of $\mathbf{a}$ ($\mathbf{A}$), respectively. For a length-n vector $\mathbf{a}$ and an integer $p \geq 1$, $\|\mathbf{a}\|_p$ denotes the p-norm of $\mathbf{a}$, i.e., $\|\mathbf{a}\|_p = \left(\sum_{i=1}^{n} |a_i|^p\right)^{1/p}$, where a_i is the ith component of $\mathbf{a}$. The 0-norm of $\mathbf{a}$ $\|\mathbf{a}\|_0$ is defined as the number of nonzero elements in $\mathbf{a}$. $\mathrm{tr}(\mathbf{A})$ denotes the trace of $\mathbf{A}$. $\mathbf{A}^{-1}$ and $\mathbf{A}^{+} = \mathbf{A}^H(\mathbf{A}\mathbf{A}^H)^{-1}$ are the inverse matrix and the pseudo-inverse matrix of $\mathbf{A}$, respectively. $\mathbf{I}_n$ is the $n \times n$ identity matrix. $(\mathbf{A})_{\mathcal{I}}$ ($(\mathbf{a})_{\mathcal{I}}$) is a submatrix (subvector) formed by collecting the columns (elements) of $\mathbf{A}$ ($(\mathbf{a})$) which correspond to indices in the set $\mathcal{I}$, and $(\mathbf{A})_{\mathcal{I}}^T$ ($(\mathbf{A})_{\mathcal{I}}^H$) denotes its transpose (conjugate transpose). For set $\mathcal{S}$, $|\mathcal{S}|$ denotes the cardinality of $\mathcal{S}$, and $\mathcal{S}^c$ is the complement of set $\mathcal{S}$. $\mathbb{E}[\cdot]$ stands for the expectation operator. $\mathcal{CN}(\mu, \Sigma)$ represents a complex Gaussian distribution with a mean μ and covariance matrix Σ.

II. System Model

A single-user massive MIMO system with M ($M \gg 1$) transmit antennas and a single receive antenna is considered. The received signal at the nth symbol time is given by

$$y[n] = \sqrt{\rho}\mathbf{h}^H[n]\mathbf{x}[n] + z[n], \tag{1}$$

where $\mathbf{h}[n] \in \mathbb{C}^{M\times 1}$ is the channel vector at the nth symbol time, $\mathbf{x}[n] \in \mathbb{C}^{M\times 1}$ is the transmit signal vector at the nth symbol time with $\mathbb{E}[\|\mathbf{x}[n]\|_2^2] = 1$, ρ is the signal-to-noise ratio (SNR), and $z[n] \in \mathbb{C}$ is the additive white noise at the nth symbol time with the distribution of $\mathcal{CN}(0, 1)$.

A. Channel Model

A block-fading model is assumed, where the channel remains constant within a fading block of L consecutive channel uses and channels of different blocks are uncorrelated. The fading model can be expressed as

$$\mathbf{h}_i = \mathbf{h}\left[(i-1)L + l\right], \quad l = 1, 2, \cdots, L, \tag{2}$$

where $\mathbf{h}_i$ denotes the channel vector of the ith block and

$$\mathbb{E}\left[\mathbf{h}_i^H\mathbf{h}_j\right] = 0, \quad i \neq j. \tag{3}$$

The channel vector can be represented in the angular domain with a proper transformation. Let $\mathbf{s}_i \in \mathbb{C}^{M\times 1}$ be the angular domain representation of channel vector $\mathbf{h}_i$. Then, the channel can be written as

$$\mathbf{h}_i = \Psi\mathbf{s}_i, \tag{4}$$

where $\Psi \in \mathbb{C}^{M\times M}$ is a unitary matrix representing the angular domain transformation [18], [28]. Since the transformation matrix is determined by the structure of the antenna array, we assume that Ψ is fixed and known. If the antennas are placed in a uniform linear array (ULA), for example, the transformation is given as

$$\Psi_{\mathrm{ULA}} = \mathbf{F}_M, \tag{5}$$

where $\mathbf{F}_M$ is the $M \times M$ discrete Fourier transform (DFT) matrix whose elements are $[\mathbf{F}_M]_{m,n} = \frac{e^{-j2\pi mn/M}}{\sqrt{M}}$. If the antennas form an $M_v \times M_h$ uniform planar array (UPA), the two-dimensional DFT (2D-DFT) can be used for the angular domain representation, i.e.,

$$\Psi_{\mathrm{UPA}} = \mathbf{F}_{M_v} \otimes \mathbf{F}_{M_h}, \tag{6}$$

where $\otimes$ denotes the Kronecker product.

Due to the compact deployment of antennas and limited scattering environment, massive MIMO channels tend to be spatially correlated and have fewer degrees of freedom than the number of antennas [3]. In the angular domain, this phenomenon is manifested as a substantial number of elements in $\mathbf{s}_i$ whose values are either zero or close to zero. Considering this fact, we assume that $\mathbf{s}_i$ has at most $k < M$ effective elements and the other elements have negligible values. In other words, a linear combination of columns of Ψ corresponding to the k largest coefficients can approximate the channel with negligible error. Let $\Omega_i \subset \{1, 2, \ldots, M\}$ be the support set of $\mathbf{s}_i$, i.e., the set of indices corresponding to the dominant elements, which we assume satisfy $|\Omega_i| = k$. Moreover, we denote the set of indices with negligible values by the complement of the support $\Omega_i^c = \{1, 2, \ldots, M\} \setminus \Omega_i$. Then the assumption is expressed as

$$\mathbf{h}_i \approx \sum_{j \in \Omega_i} \mathbf{s}_i(j) \boldsymbol{\psi}_j, \tag{7}$$

where $\mathbf{s}_i(j)$ is the jth element of the vector $\mathbf{s}_i$ and $\boldsymbol{\psi}_j$ is the jth column of Ψ. In compressed sensing theory, $\mathbf{s}_i$ is said to be *approximately sparse* if k is much smaller than M, i.e., $k \ll M$ [12].

In practical massive MIMO systems, the support of $\mathbf{s}_i$ is determined by the geometry of the receiver and scattering environment [18], [19]. Reflecting the observation that the environment may change slowly in realistic deployment scenarios and can be modeled as an evolution of scattering clusters [27], it is assumed that the supports of consecutive fading blocks are similar to each other (i.e., Ω_{i-1} and Ω_i share many common entries). Let k_s be the maximum number of support elements that are newly added to Ω_i from those not included in Ω_{i-1}, i.e.,

$$|\Omega_i \setminus \Omega_{i-1}| \leq k_s. \tag{8}$$

This means that at most k_s elements are newly added to the support of $\mathbf{s}_i$ from those not included in the support of $\mathbf{s}_{i-1}$. Then, the slow rate of change of the support can be expressed as

$$k_s \ll |\Omega_{i-1}^c|. \tag{9}$$

As a result, most of the elements corresponding to the previous support Ω_{i-1} are nonzero while most of the elements corresponding to the complement of previous support Ω_{i-1}^c are zero. In other words, if we separate $\mathbf{s}_i$ into two vectors based on the previous support as

$$\mathbf{s}_{i,d} = (\mathbf{I}_M)_{\Omega_{i-1}}^T \mathbf{s}_i, \tag{10}$$

and

$$\mathbf{s}_{i,s} = (\mathbf{I}_M)_{\Omega_{i-1}^c}^T \mathbf{s}_i, \tag{11}$$

where $\mathbf{I}_M$ denotes the $M \times M$ identity matrix, then $\mathbf{s}_{i,d} \in \mathbb{C}^k$ is a dense vector, while $\mathbf{s}_{i,s} \in \mathbb{C}^{M-k}$ is a sparse vector. Let $\Lambda_{i,d} \subset \{1, 2, \ldots, k\}$ and $\Lambda_{i,s} \subset \{1, 2, \ldots, M-k\}$ be the index sets corresponding to the nonzero elements of $\mathbf{s}_{i,d}$ and $\mathbf{s}_{i,s}$, respectively, with cardinalities $|\Lambda_{i,d}| \geq k - k_s$ and $|\Lambda_{i,s}| \leq k_s$. Throughout the paper, we assume that the channel sparsity parameters k and k_s are known. In practical systems, these values need to be chosen with the consideration of long term statistics.

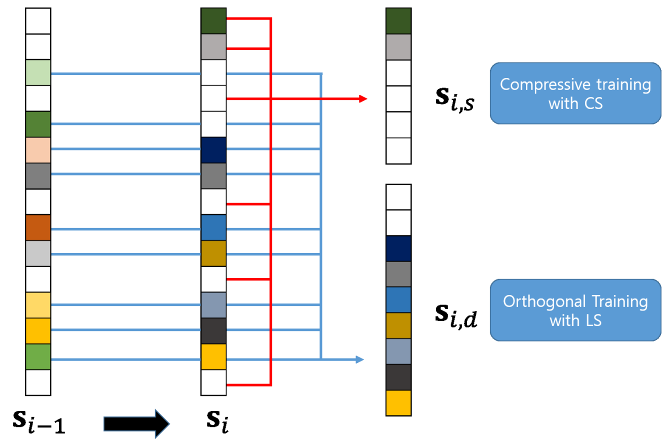

Fig. 1. Illustration of the channel model and the concept of the proposed channel estimation scheme.

The channel model is depicted in Fig. 1, where the white and colored elements of vectors denote the zero and nonzero elements of the vectors, respectively. Although the values of the coefficients (colors) can vary independently, the supports of $\mathbf{s}_{i-1}$ and $\mathbf{s}_i$ have only two different elements. As a result, $\mathbf{s}_i$ can be decomposed into a dense vector $\mathbf{s}_{i,d}$ and a sparse vector $\mathbf{s}_{i,s}$ exploiting the previous support Ω_{i-1}.

B. Downlink Channel Estimation

To achieve the high spectral efficiency available in massive MIMO systems, coherent transmission exploiting CSIT is required. In FDD systems, CSIT can be achieved through downlink training and uplink feedback. If we neglect the feedback delay, a channel block can be divided into a channel training period of length T_p and a data transmission period of length $L - T_p$ with a coherence block of length L, as depicted in Fig. 2. Since the length of a block is limited by the channel coherence time, the available number of channel uses for downlink transmission decreases as the training time T_p increases. Hence, an efficient channel training scheme that utilizes a small T_p needs to be developed for massive MIMO systems.

The received signal vector during the training period can be expressed as

$$\begin{aligned} \mathbf{y}_{i,T}^T &= \left[y[(i-1)L+1], \ldots, y[(i-1)L+T_p]\right] \\ &= \sqrt{\rho_T}\mathbf{h}_i^H \mathbf{X}_T + \mathbf{z}_{i,T}^T, \end{aligned} \tag{12}$$

where $\mathbf{X}_T = \left[\mathbf{x}_T[1], \ldots, \mathbf{x}_T[T_p]\right] \in \mathbb{C}^{M \times T_p}$ is a pilot sequence with the constraint of $\mathrm{tr}(\mathbf{X}_T^H \mathbf{X}_T) = T_p$, ρ_T is the SNR during the training period, and $\mathbf{z}_{i,T} = \left[z[(i-1)L+1], \ldots, z[(i-1)L+T_p]\right]^T$ is a noise vector. Using the received vector $\mathbf{y}_{i,T}$, the receiver estimates the channel and sends it to the transmitter through a feedback link.

In the absence of knowledge of the channel statistics, the conventional LS approach estimates the channel as

$$\hat{\mathbf{h}}_i^{LS} = \frac{1}{\sqrt{\rho_T}} \left(\mathbf{y}_{i,T}^T \mathbf{X}_T^+\right)^H, \tag{13}$$

where $\mathbf{X}_T^+$ is the pseudo-inverse of $\mathbf{X}_T$. However, this approach requires $T_p \geq M$ and causes training to consume

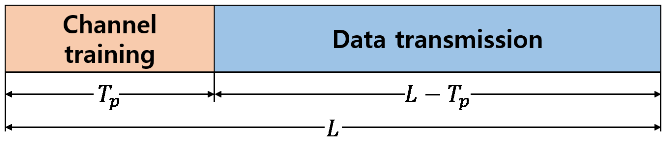

Fig. 2. Downlink channel block.

a large amount of the downlink resources in massive MIMO systems.

If the transmitter and the receiver know the channel statistics and the support of $\mathbf{s}_i$ a priori, the channel can be estimated using a shorter training sequence. Specifically, a sequence of length T_p satisfying $T_p \geq k$ is enough to train the coefficients $\mathbf{s}_i(j)$ with $j \in \mathbf{\Omega}_i$. The training sequence is constructed as

$$\mathbf{X}_T = \Psi_{\Omega_i} \tilde{\mathbf{X}}_T, \tag{14}$$

where Ψ_{Ω_i} is the $M \times k$ matrix obtained by collecting columns of Ψ corresponding to the indices in $\mathbf{\Omega}_i$, and $\tilde{\mathbf{X}}_T \in \mathbb{C}^{k \times T_p}$ is a pilot sequence for training the k nonzero coefficients. Given $T_p \geq k$, a conventional estimation scheme such as LS can estimate the coefficients.

Since the support varies according to the relative location of the receiver and the evolution of the scattering environment, it is difficult to know the exact support of the channel in practical systems. Without knowledge of the support, in general, channel estimation using a pilot of length $T_p < M$ is an underdetermined problem. If $\mathbf{s}_i$ is sparse (or approximately sparse), i.e., $k \ll M$, CS can be utilized for the channel estimation. For the CS approach, the pilot is designed using

$$\mathbf{X}_{CS} = \Psi \Phi_{CS}^H, \tag{15}$$

where $\Phi_{CS} \in \mathbb{C}^{T_p \times M}$ is a compressive measurement matrix that satisfies a certain condition. After receiving the training signals, the receiver can estimate the channel using a sparse recovery algorithm such as basis pursuit (BP) or orthogonal matching pursuit (OMP). Much work has been devoted to characterizing the conditions that a measurement matrix Φ_{CS} must have to guarantee the successful support recovery in terms of the restricted isometry property (RIP) [14] and mutual coherence [15]. In the literature, it can be found that $T_p = O(k \log(M/k))$ measurements are required to recover a k-sparse channel, and $T_p \approx 4k$ is usually accepted in many practical applications [12]. Although these results provide great opportunities to reduce the length of training sequence, massive MIMO channels are hardly sparse enough to directly apply CS unless a millimeter wave band is used.

III. CS-Aided Channel Training

The slowly-varying support model, described in the previous section, can be exploited to reduce the training overhead. For the overhead reduction, we introduce a support tracking concept whose operation is described in Fig. 3. While at least M channel uses are required at the initial stage where no prior knowledge is available, the succeeding block channel can be trained with a reduced overhead by utilizing the support extracted from the previous estimation. Note that a length M training sequence can be used initially or when a complete tracking failure of the support occurs.

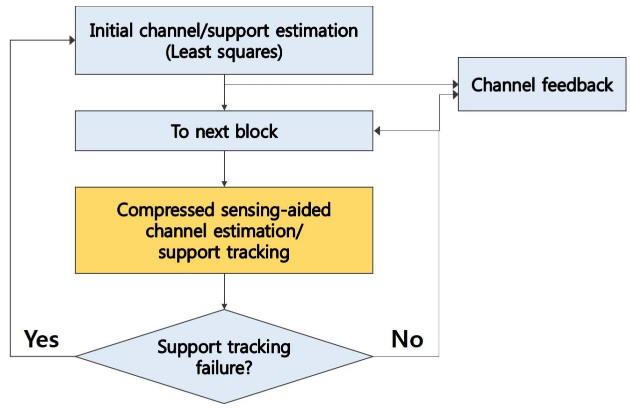

Fig. 3. Pilot overhead reduction exploiting slowly-varying support.

The key component for the proposed framework is the estimation scheme exploiting the previous support information. To reduce the training overhead, we propose a CS-aided channel training scheme that utilizes the LS and CS approaches simultaneously. Based on the observation that the support of the transformed channel vector changes very slowly, the proposed scheme separates the channel into two vectors based on the previous support and estimates them with different strategies. In the following, we assume that the channel sparsity parameters k and k_s are known. In practice, k and k_s can be obtained by observing long-term estimations [29], or a learning-based algorithm can be utilized to adapt the values to the channel variation.

A. Training Sequence Design

To reduce the training overhead, the proposed scheme separates the angular domain channel based on the previous support and pursues different strategies for the estimation of separated channels. For the channel separation, the training sequence is divided into two parts. While a part of length k is dedicated for the sounding of coefficients of $\mathbf{s}_{i,d}$ defined in (10), the remaining part of length $T_p - k$ is utilized for the estimation of $\mathbf{s}_{i,s}$ defined in (11).[1] The pilot structure can be expressed as

$$\mathbf{X}_{i,T} = \begin{bmatrix} \mathbf{X}_{i,d} & \mathbf{X}_{i,s} \end{bmatrix}, \tag{16}$$

where $\mathbf{X}_{i,d} \in \mathbb{C}^{M \times k}$ is the training sequence for $\mathbf{s}_{i,d}$, and $\mathbf{X}_{i,s} \in \mathbb{C}^{M \times (T_p - k)}$ is for $\mathbf{s}_{i,s}$.

To estimate $\mathbf{s}_{i,d}$, the former part $\mathbf{X}_{i,d}$ is designed as

$$\mathbf{X}_{i,d} = \Psi_{\Omega_{i-1}} \tilde{\mathbf{X}}_d, \tag{17}$$

where $\tilde{\mathbf{X}}_d \in \mathbb{C}^{k \times k}$ is an orthonormal matrix satisfying $\tilde{\mathbf{X}}_d^H \tilde{\mathbf{X}}_d = \mathbf{I}_k$. With this construction, the dense vector of length k can be estimated through the LS filtering since we have k measurements.

The latter part $\mathbf{X}_{i,s}$ is used for the estimation of $\mathbf{s}_{i,s}$. Since $\mathbf{s}_{i,s}$ is sparse according to the slowly-varying support model, we can estimate it with a compressed training sequence whose length is $T_p - k < |\Omega_{i-1}^c| = M - k$. To estimate the sparse

[1] $T_p \geq k$ is assumed since channel estimation with a pilot of length $T_p < k$ is underdetermined even with the knowledge of the instantaneous support.

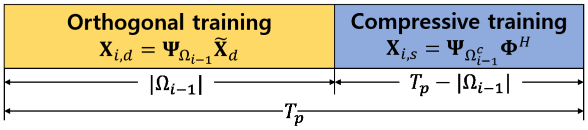

Fig. 4. Pilot structure for the proposed channel estimation.

vector $\mathbf{s}_{i,s}$ with the CS approach, the training sequence is constructed as

$$\mathbf{X}_{i,s} = \Psi_{\Omega_{i-1}^c}\Phi^H, \tag{18}$$

where $\Phi \in \mathbb{C}^{(T_p-k)\times(M-k)}$ is a measurement matrix that satisfies the conditions for successful recovery.[2] The pilot structure for the channel separation is illustrated in Fig. 4.

B. Channel Estimation

With the proposed training sequence design, the received signal at the training phase is given by

$$\mathbf{y}_{i,T}^T = \begin{bmatrix}\mathbf{y}_{i,d}^T & \mathbf{y}_{i,s}^T\end{bmatrix}, \tag{19}$$

where

$$\begin{aligned}\mathbf{y}_{i,d}^T &= \sqrt{\rho_T}\mathbf{h}_i^H\mathbf{X}_{i,d} + \mathbf{z}_{i,d}^T\\ &= \sqrt{\rho_T}\mathbf{s}_i^H(\mathbf{I}_M)_{\Omega_{i-1}}\tilde{\mathbf{X}}_d + \mathbf{z}_{i,d}^T\\ &= \sqrt{\rho_T}\mathbf{s}_{i,d}^H\tilde{\mathbf{X}}_d + \mathbf{z}_{i,d}^T,\end{aligned} \tag{20}$$

and

$$\begin{aligned}\mathbf{y}_{i,s}^T &= \sqrt{\rho_T}\mathbf{h}_i^H\mathbf{X}_{i,s} + \mathbf{z}_{i,s}^T\\ &= \sqrt{\rho_T}\mathbf{s}_i^H(\mathbf{I}_M)_{\Omega_{i-1}^c}\Phi^H + \mathbf{z}_{i,s}^T\\ &= \sqrt{\rho_T}\mathbf{s}_{i,s}^H\Phi^H + \mathbf{z}_{i,s}^T.\end{aligned} \tag{21}$$

The vector $\mathbf{s}_{i,d}$ can be estimated from $\mathbf{y}_{i,d}$. With the LS approach, the estimate of $\mathbf{s}_{i,d}$ is obtained as

$$\tilde{\mathbf{s}}_{i,d} = \frac{1}{\sqrt{\rho_T}}\tilde{\mathbf{X}}_d\mathbf{y}_{i,d}^* = \mathbf{s}_{i,d} + \frac{1}{\sqrt{\rho_T}}\tilde{\mathbf{X}}_d\mathbf{z}_{i,d}^*, \tag{22}$$

where $\tilde{\mathbf{s}}_{i,d}$ denotes the estimate of $\mathbf{s}_{i,d}$. On the other hand, $\tilde{\mathbf{s}}_{i,s}$, the estimate of $\mathbf{s}_{i,s}$, can be obtained from $\mathbf{y}_{i,s}$ using a sparse recovery algorithm. Among various algorithms, we focus on the OMP algorithm [30] whose procedures are as follows:

1) Initialize the residual $\mathbf{r}_0 = \mathbf{y}_{i,s}^*$, the index set $\Lambda_0 = \emptyset$, and the iteration counter $t = 1$.
2) Find the index of the column of Φ which has the maximum correlation with the residual $\mathbf{r}_{t-1}$, i.e., $j_t = \underset{j\in\{1,2,\ldots,M-k\}}{\operatorname{argmax}} |\phi_j^H\mathbf{r}_{t-1}|$, where ϕ_j denotes the jth column of the measurement matrix Φ.
3) Update the index set $\Lambda_t = \Lambda_{t-1}\cup\{j_t\}$.
4) Estimate the signal based on the current index set $\mathbf{u}_t = (\mathbf{I}_{M-k})_{\Lambda_t}\left(\Phi_{\Lambda_t}^H\Phi_{\Lambda_t}\right)^{-1}\Phi_{\Lambda_t}^H\mathbf{y}_{i,s}^*$ and update the residual $\mathbf{r}_t = \mathbf{y}_{i,s}^* - \Phi\mathbf{u}_t$.
5) If $t = k_s$, stop and estimate $\tilde{\mathbf{s}}_{i,s} = \frac{1}{\sqrt{\rho_T}}\mathbf{u}_t$. Otherwise, update $t = t + 1$ and return to Step 2.

Note that if the magnitude of the residual signal after k_s iterations is larger than a predefined threshold, we consider it as a support tracking failure and perform an initial channel estimation with a long sequence, as described in Fig. 3.

With the estimates $\tilde{\mathbf{s}}_{i,d}$ and $\tilde{\mathbf{s}}_{i,s}$, a primary estimate can be obtained as

$$\tilde{\mathbf{s}}_i = (\mathbf{I}_M)_{\Omega_{i-1}}\tilde{\mathbf{s}}_{i,d} + (\mathbf{I}_M)_{\Omega_{i-1}^c}\tilde{\mathbf{s}}_{i,s}. \tag{23}$$

Note that $\tilde{\mathbf{s}}_i$ can have up to $k+k_s$ nonzero elements due to the nature of the OMP process, while $\mathbf{s}_i$ has at most k nonzero elements. To refine the estimate and obtain the support set for the estimation of the next block channel, we construct a final estimate of the channel by selecting indices corresponding to the k largest absolute values of $\hat{\mathbf{s}}_i$, i.e.,

$$\hat{\mathbf{s}}_i = \underset{\mathbf{s}\in\mathbb{C}^{M\times 1},\|\mathbf{s}\|_0\le k}{\operatorname{argmin}} \|\mathbf{s} - \tilde{\mathbf{s}}_i\|_1. \tag{24}$$

When OMP is used for support recovery in the presence of noise, the stopping rule in Step 5 generally includes the norm of the residual, i.e., stop if $t = k_s$ or $\|\mathbf{r}_t\|_2 < \eta$, where η is threshold determined by the noise level [31], [32].[3] In the proposed scheme, however, the iteration continues until $t = k_s$. Instead of thresholding, the support is pruned as (24) after combining the results of LS and CS estimation.

C. Estimation Error

In this subsection, the estimation error of the proposed scheme is analyzed in terms of the mean squared error (MSE), which is defined as

$$MSE = \mathbb{E}\left[\|\mathbf{e}_i\|_2^2\right] = \mathbb{E}\left[\|\mathbf{h}_i - \hat{\mathbf{h}}_i\|_2^2\right], \tag{25}$$

where $\mathbf{e}_i = \mathbf{h}_i - \hat{\mathbf{h}}_i$ denotes the estimation error vector. For analytical convenience, we assume that supports of consecutive channel blocks have exactly k_s different elements, i.e., $|\Omega_i \setminus \Omega_{i-1}| = |\Omega_{i-1} \setminus \Omega_i| = k_s$. Furthermore, we assume that the OMP algorithm is used for sparse recovery, and the measurement matrix Φ is characterized in terms of the RIP, which is defined as below.

Definition 1 (Restricted Isometry Property [33]): A matrix $\mathbf{A} \in \mathbb{C}^{m\times n}$ $(m < n)$ is said to satisfy the restricted isometry property (RIP) of order K with an isometry constant $\delta_K \in (0, 1)$ if

$$(1-\delta_K)\|\mathbf{v}\|_2^2 \le \|\mathbf{A}\mathbf{v}\|_2^2 \le (1+\delta_K)\|\mathbf{v}\|_2^2 \tag{26}$$

holds for all $\mathbf{v} \in \mathbb{C}^{n\times 1}$ such that $\|\mathbf{v}\|_0 \le K$, where $\|\mathbf{v}\|_0$ denotes the number of nonzero elements of $\mathbf{v}$.

By leveraging the proposed estimation procedure and the sufficient condition for the support recovery, which has been developed in [32], we obtain the following results about the MSE.

[2]The sufficient condition of measurement matrix which guarantees the support recovery can be characterized in terms of various criteria. Among them, the RIP condition is adopted in this paper. Readers who are interested in other conditions are referred to [13]–[15] and references therein.

[3]This stop criterion, along with the pilot structure, makes it difficult to generalize the proposed scheme for multi-user multi-cell systems, where interference power is time-varying. The extension to general multi-user systems remains for further studies.

Proposition 1: Assume that the SNR ρ_T is sufficiently large. When the OMP algorithm is utilized and the scaled measurement matrix $\bar{\Phi} = \sqrt{\frac{M-k}{T_p-k}}\Phi$ satisfies the RIP of order k_s+1 with isometry constant $\delta_{k_s+1} < \frac{1}{\sqrt{k_s}+1}$, the induced MSE is given as

$$MSE = \frac{k-k_s}{\rho_T} + \frac{1}{\rho_T}\mathrm{tr}\left(\left(\Phi_{\Lambda_{i,s}}^H \Phi_{\Lambda_{i,s}}\right)^{-1}\right), \tag{27}$$

where $\Lambda_{i,s} \subset \{1, 2, \ldots, M-k\}$ is the support of $\mathbf{s}_{i,s}$. The MSE is bounded as

$$\frac{1}{\rho_T}\left(k-k_s+\frac{(M-k)k_s}{(T_p-k)(1+\delta_{k_s+1})}\right) \leq MSE \leq \frac{1}{\rho_T}\left(k-k_s+\frac{(M-k)k_s}{(T_p-k)(1-\delta_{k_s+1})}\right). \tag{28}$$

Proof: See Appendix. □

From the result, we can see that the upper bound in (28) increases as k_s increases for a fixed $T_p \leq M$ since the isometry constant of a matrix increases with its order. This implies that as the consecutive channels share larger part of support (i.e., smaller k_s), the proposed scheme can estimate the channel with smaller error, validating the efficient use of the prior support information.

IV. Discussions

A. Design of Measurement Matrix

As seen in the previous section, the estimation performance of the proposed scheme highly depends on the design of the measurement matrix Φ. Specifically, the support can be recovered with higher probability and estimation error becomes smaller as the isometry constant decreases. For these reasons, the design of a matrix satisfying the RIP with a small isometry constant has been of great interest in various CS applications [12]. Since finding a matrix with the smallest isometry constant is infeasible due to the combinatorial nature of the RIP [34], most CS research has considered randomly generated sensing matrices. For example, it has been shown that a matrix with i.i.d. random entries drawn from a Gaussian distribution [35] or a matrix designed by collecting random rows of a unitary matrix (e.g., a DFT matrix) can satisfy the RIP with a high probability. Specifically, it is known that a Gaussian matrix $\mathbf{A} \in \mathbb{C}^{m\times n}$ satisfies RIP of order K with an overwhelming probability if $m \geq C \cdot \log(n/K)$, where C is a constant depending on the isometry constant δ_K [12]. Based on the result, we can construct a pilot signal according to the process of (16)-(18) with a randomly generated measurement matrix Φ. On the other hand, an optimization approach for measurement matrix design is also being actively studied by establishing objective functions and developing corresponding algorithms [36], [37]. Although there is no performance measure that is directly related to the CS performance, adopting a measurement matrix developed by these algorithms as Φ can improve the estimation performance.

B. Extension to MIMO Systems

Throughout the paper, we have considered MISO systems where the receiver is equipped with a single antenna. When multiple antennas are deployed at the receiver, the input-output relation of (1) is replaced by

$$\mathbf{y}[n] = \sqrt{\rho}\mathbf{H}_i^H\mathbf{x}[n] + \mathbf{z}[n], \tag{29}$$

where $\mathbf{H}_i \in \mathbb{C}^{M\times N}$ denotes the channel matrix between the transmitter and the receiver, N is the number of receive antennas, and $\mathbf{z}[n] \in \mathbb{C}^{N\times 1}$ is the Gaussian noise vector. In the MIMO setting, the channel estimation is the problem of estimating $\mathbf{H}$ with the received training signal

$$\mathbf{Y}_{i,T}^T = \sqrt{\rho_T}\mathbf{H}_i^H\mathbf{X}_T + \mathbf{Z}_{i,T}^T. \tag{30}$$

While the estimation of $\mathbf{H}_i$ can be performed column-wise operation, the estimation performance can be improved by exploiting the joint sparsity of MIMO channels. With a transformation similar to the MISO case, the channel matrix can be represented in the angular domain [21], i.e.,

$$\mathbf{H}_i = \Psi_t\mathbf{S}_i\Psi_r^H, \tag{31}$$

where $\Psi_t \in \mathbb{C}^{M\times M}$ and $\Psi_r \in \mathbb{C}^{N\times N}$ denote the angular transform matrices at the transmitter and the receiver, respectively. Due to the relatively rich scattering experienced by the receiver, the columns of $\mathbf{S}_i$ usually have the same support [21], [38]. Hence, the proposed pilot design method, which exploits the support information of the previous block, is effective for systems with multiple receive antennas. Moreover, the joint sparsity of the columns can be utilized to enhance the estimation accuracy. The recovery of multiple sparse vectors with an identical support has been extensively studied [39], [40], and it was shown that the recovery performance can be improved by exploiting the joint sparsity. Consequently, the performance of the proposed scheme can be further improved in MIMO systems by utilizing a recovery algorithm developed for simultaneous sparse signal, which will be seen in Sec V.

C. Comparison to CS With Partial Support Information

Since a slowly-varying support can be found in various applications where sequential sparse signals are reconstructed, attempts to exploit the partially known support have naturally arisen. Particularly, several CS algorithms that utilize partially known support information in sparse recovery have recently been introduced [29], [41]–[44]. In [41], a simple approach that applying CS on the LS residual computed using the previous support was proposed as a first CS solution utilizing partial support information. In [42], another solution, referred to as *modified-CS*, was proposed by the same authors, which finds a support that contains the smallest number of additional elements to the prior information. The modified-CS was further developed into *weighted ℓ_1 minimization* in [43] by taking into account the expected values of nonzero elements. In [29], a greedy pursuit-based approach, referred to as *modified subspace pursuit (M-SP)*, was proposed to incorporate the partial support information. An approximate message passing algorithm [45] was also modified to exploit the partially known support [44]. Moreover, M-SP [29] and weighted ℓ_1 minimization [46] have been applied to the estimation of massive MIMO channels with a motivation similar to our work.

Hence, it is worth comparing the proposed channel estimation scheme with existing CS algorithms which incorporate prior support information. The key difference is that the proposed scheme utilizes prior support information to construct the pilot signal, while existing algorithms utilize the information only during their recovery process. While various CS algorithms have been modified to incorporate the partial support information in the recovery process, none of the previous work considers the adaptation of the measurement matrix to the best of our knowledge. Since the proposed scheme adapts the pilot sequence, it can estimate channels with more relaxed support condition and shows better estimation performance. The superior performance of the proposed scheme is numerically verified in Sec. V. Moreover, the proposed scheme can be combined with various sparse recovery algorithms such as convex optimization or message passing algorithms to enhance the performance. However, due to the difference, the application of the proposed scheme is restricted to systems with a single user or multiple users with a common support while the existing recovery algorithms can be applied to a multi-user massive MIMO channel estimation.

V. Simulation Results

In this section, the performance of the proposed scheme is compared with several estimation schemes. For the proposed scheme, the measurement matrix Φ is randomly drawn from the i.i.d. Gaussian distribution and normalized to satisfy the power constraint $\mathrm{tr}\left(\Phi\Phi^H\right) = T_p - k$. As a sparse recovery algorithms, OMP is used. The baseline schemes are as follows.

- **Genie-aided LS:** The support $\boldsymbol{\Omega}_i$ is assumed to be known. Using the information, the pilot is designed to train the subspace the support spans, and LS estimation is used. This ideal scheme provides a lower bound for MSE.
- **Static LS:** Only the support of the first block is known. The pilot is designed using the outdated support $\boldsymbol{\Omega}_1$ for all time and LS is used for estimation.
- **Random LS:** A randomly generated orthonormal sequence of length T_p is used for training and LS is used as an estimation filter.
- **OMP:** A random Gaussian matrix of size $M \times T_p$ is used for training signal. The OMP algorithm recovers the channel.
- **M-SP:** A random Gaussian matrix of size $M \times T_p$ is used for training signal. The M-SP algorithm, introduced in [29], recovers the channel incorporating the previous support information. Similarly to the proposed scheme, M-SP successively updates the support information.

First, we consider a toy channel model. In this model, an angular channel vector is assumed to have exactly k nonzero elements, while the remaining $M - k$ elements are zero. The support is randomly selected such that $k - k_s$ indices are preserved from the previous support. The nonzero coefficients are drawn from an i.i.d. complex Gaussian distribution with zero mean and unit variance. For the simulation, we set $M = 100$, $k = 40$, and $k_s = 3$.

Fig. 5 and Fig. 6 show the normalized mean squared error (NMSE) $\frac{\mathbb{E}\|\mathbf{h}_i - \hat{\mathbf{h}}_i\|_2^2]}{\mathbb{E}[\|\mathbf{h}_i\|_2^2}$ and the normalized beamforming gain $\mathbb{E}\left[\frac{|\mathbf{h}_i^H\hat{\mathbf{h}}_i|^2}{\|\mathbf{h}_i\|_2\|\hat{\mathbf{h}}_i\|_2}\right]$ of various schemes, respectively, where the training is performed with the SNR of $\rho_T = 20$ [dB], and the length of training sequence varies from k to M. As expected, the low degrees of freedom of the channel cannot be exploited using the random orthogonal pilot. OMP also fails to estimate with a reduced pilot overhead. This implies that the sparsity level of the channel (40 of 100 elements are nonzero) is not enough to be recovered by the CS approach. Since the support changes slowly, a static approach shows a smaller error compared to the random approaches such as orthogonal sounding and OMP. However, the coefficients corresponding to the changed k_s support elements cannot be estimated with the static approach. M-SP, which exploits the previous support information in the recovery process, shows better performance than the conventional OMP. The proposed CS-aided channel estimation, which can estimate the sparse vector corresponding to the complement of the previous support by adding a short training sequence (about 20 additional symbols), outperforms the M-SP and shows a performance close to the ideal Genie-aided scheme. This results show the effectiveness of the channel separation and partial application of a CS technique.

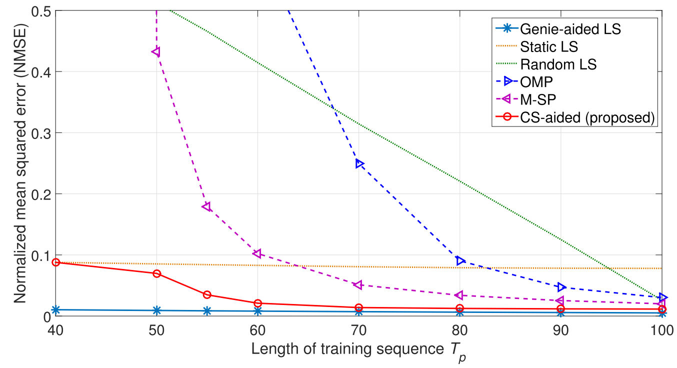

Fig. 5. NMSE versus the length of the training sequence with $M = 100$, $k = 40$, $k_s = 3$, and SNR= 20 dB.

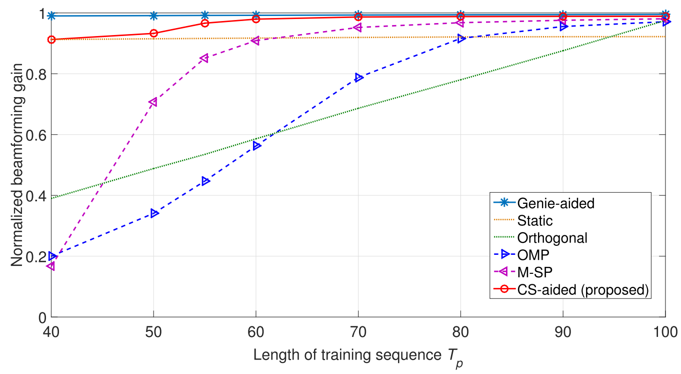

Fig. 6. Normalized beamforming gain versus the length of the training sequence with $M = 100$, $k = 40$, $k_s = 3$, and (training) SNR= 20 dB.

Fig. 7 shows the NMSE of the estimation schemes when the training is performed with varying SNR. The length of the training sequence is set to $T_p = 60$. Since the CS algorithms is inherently sensitive to noise level, the estimation schemes that utilize CS, including the proposed scheme, show poor performance in low SNR region. On the other hand,

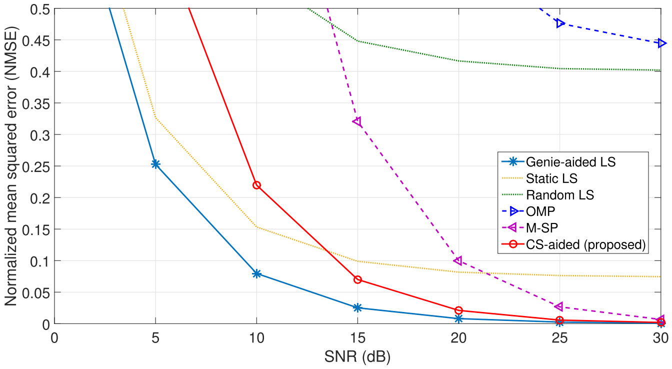

Fig. 7. NMSE versus SNR with $M = 100$, $k = 40$, $k_s = 3$, and $T_p = 60$.

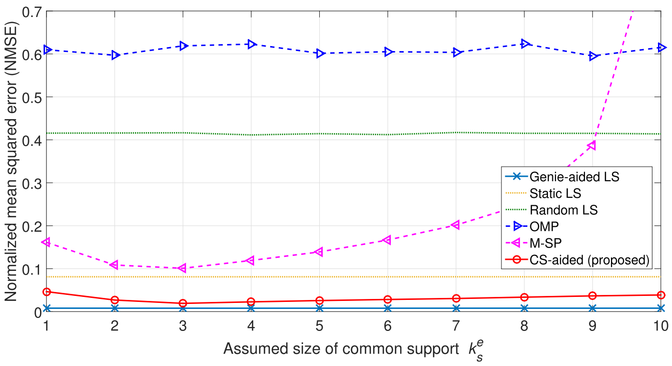

Fig. 8. NMSE versus the mismatched parameter k_s^e with $M = 100$, $k = 40$, $k_s = 3$, and SNR= 20 dB.

the LS based schemes (Genie-aided and Static) show relatively low error. With an SNR larger than 15 dB, the proposed scheme outperforms the Static LS and shows reasonable performance. In all region, the proposed scheme shows lower error than the conventional CS algorithms.

Thus far, we assumed that the channel evolution parameter k_s is exactly known. In practical implementations, the assumption may be unrealistic. Fig. 8 shows the NMSE performance when wrong value of k_s^e is known. We consider mismatched parameters $1 \le k_s^e \le 10$, while the real channel evolves with $k_s = 3$. As expected, estimation schemes that exploit the previous support information (M-SP and CS-aided) show performance degradation when the parameter mismatch occurs. However, the degradation of the proposed scheme is marginal, while M-SP is highly vulnerable to the mismatch. The robustness of the proposed scheme can be explained with two reasons. First, since only a partial part of the channel is estimated with a CS algorithm, the impact of the mismatch is limited. Moreover, even though wrong indices are selected by the CS algorithm operating with $k_s^e > k_s$, the pruning step can eliminate the effect of wrong indices.

In the following, the estimation performance is evaluated in a realistic channel evolution scenario. A ray-based channel model for uniform array antenna configuration, which is used in [47], is adopted. According to the model, a channel vector of $M_v \times M_h$ UPA is constructed as

$$\mathbf{h} = \sum_{k=1}^{L} \frac{e^{j\varphi_k}}{\sqrt{L}} \mathbf{a}_h(u_k) \otimes \mathbf{a}_v(v_k), \tag{32}$$

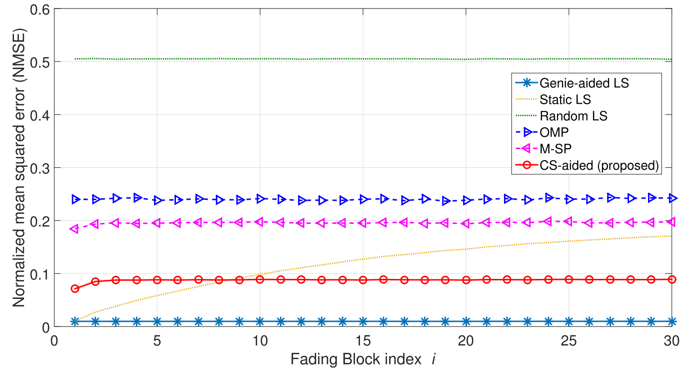

Fig. 9. NMSE behavior according to the scattering evolution with a ULA of $M = 128$, $T_p = 64$, and SNR= 20 dB.

where L is the number of independent paths, φ_k is a random phase for the kth path, $\mathbf{a}_h(u_k)$ and $\mathbf{a}_v(v_k)$ are antenna response vectors in horizontal and vertical axes, defined as

$$\mathbf{a}_h(u_k) = [1, e^{-ju_k}, \ldots, e^{-j(M_h-1)u_k}], \tag{33}$$

and

$$\mathbf{a}_v(v_k) = [1, e^{-jv_k}, \ldots, e^{-j(M_v-1)v_k}]. \tag{34}$$

u_k and v_k are defined as

$$u_k = \frac{2\pi D_h}{\lambda} \sin(\theta_k)\cos(\phi_k), \tag{35}$$

and

$$v_k = \frac{2\pi D_v}{\lambda} \cos(\theta_k), \tag{36}$$

where D_h and D_h are antenna spacing in horizontal/vertical axes, λ is the carrier wavelength, and θ_k and ϕ_k are vertical and horizontal angles of departure (AoD). The AoD pair of each path is assume to be normal distributed as $\theta_k \sim \mathcal{N}(\theta, \xi)$ and $\phi_k \sim \mathcal{N}(\phi, \sigma)$ with center angles θ, ϕ and angular spreads ξ, σ. In the simulation, we assumed $L = 15$ paths. To model the slow variation in scattering environment, we assume that one path disappears and another randomly generated path appears every fading block, while the total number of paths remains unchanged. Note that random phase φ_k is independently generated every block in this model.

Fig. 9 and Fig. 10 show the NMSE of various schemes when the channel evolves according to this model. In Fig. 9, a horizontal ULA of $M = 128$ antennas is considered with antenna spacing $D_h = \lambda/2$. The mean and standard deviation of azimuth AoD are set to $\phi = \pi/3$ and $\sigma = \pi/6$, respectively. The DFT matrix is used for the angular domain transformation. For each block, $T_p = M/2 = 64$ channel uses are allocated for training signal, and the parameters for the proposed scheme are set as $k = 32$ and $k_s = 10$. In Fig. 10, a UPA of $(M_h \times M_v) = (32 \times 8)$ is considered with antenna spacing of $D_h = \lambda/2$ and $D_v = \lambda$. The mean and standard deviation of elevation AoD are set to $\theta = \pi/4$ and $\xi = \pi/12$, respectively, while azimuth AoD of each path follows the same distribution as the ULA setting. The 2D-DFT is used for the angular domain transformation. The length of training sequence and parameters are set as $T_p = 128$, $k = 64$, and $k_s = 15$. For both scenarios, the SNR is set to $\rho_T = 20$ [dB].

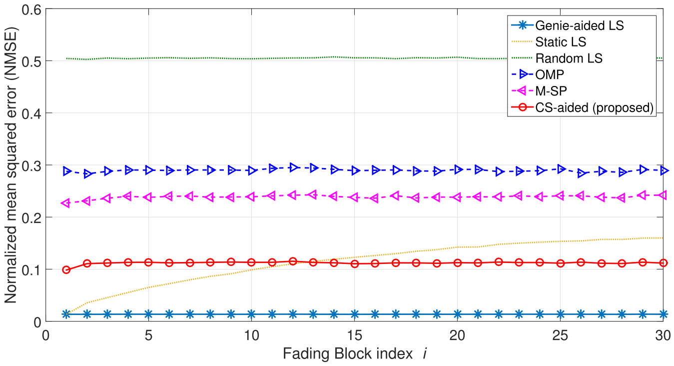

Fig. 10. NMSE behavior according to the scattering evolution with a UPA of $(M_h \times M_v) = (32 \times 8)$, $T_p = 128$, and SNR= 20 dB.

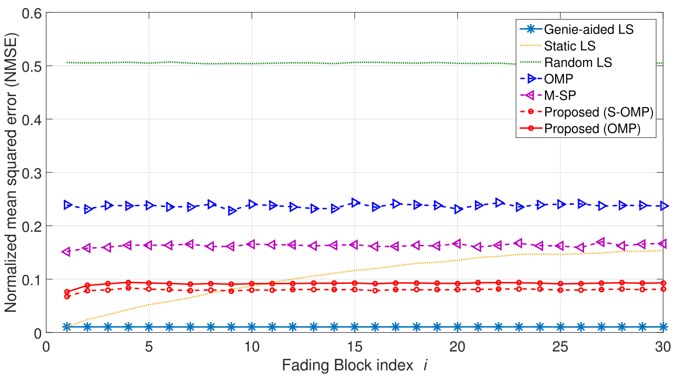

Fig. 11. NMSE behavior according to the scattering evolution with $M = 128$ transmit antennas in ULA, $N = 2$ receive anennas, $T_p = 64$, and SNR= 20 dB.

The results show that the error of the static approach utilizing the outdated support $\mathbf{\Omega}_1$ is kept small with a negligible change of the support, but increases as the scattering environment keeps changing. This implies that periodic update of channel statistics is required for the reduction of pilot, which can be a huge burden in massive MIMO systems. Moreover, both the random LS and the OMP fail to adequately estimate the channel. While M-SP, which utilizes the prior support information in its recovery process, shows a better performance compared to the conventional OMP, it still shows high estimation error. However, the proposed scheme shows stable error performance since it tracks the statistics without explicit update process. It should be noted that the performance of CS-based schemes (OMP, M-SP, and the proposed) is degraded in realistic ray-based channel environments compared to the ideal sparse channel model adopted in Fig. 5-7. This is because realistic channels are not exactly sparse in angular domain due to the limited resolution of DFT sampling, and the performance can be improved by using advanced sparse recovery techniques [48].

Fig. 11 shows the estimation performance in a MIMO system with $N = 2$ receive antennas, in which the transmit side setting is the same as the ULA of Fig. 9. The angle-of-arrival (AoA) of each path at the receiver is assumed to be uniformly distributed in $[0, 2\pi]$. While most baseline schemes show performance similar to that of the MISO scenario, the performance of M-SP is improved since the algorithm implicitly utilizes the joint sparsity. The proposed scheme was performed with two sparse recovery algorithms. Solid line shows the performance with the OMP algorithm applied to each row, and dashed line denotes the performance with simultaneous OMP (S-OMP) exploiting the joint sparsity [40]. As predicted in IV-B, the performance is slightly improved by exploiting the joint sparsity. Since the proposed scheme can incorporate various sparse recovery algorithms, it is expected that the performance can be further improved by adopting a more advanced algorithm.

VI. Conclusion

In this paper, a CS-aided channel estimation scheme, which utilizes LS and CS simultaneously, was proposed for FDD massive MIMO systems. Under the assumptions that massive MIMO channels have fewer degrees of freedom than the number of transmit antennas, and the support of the angular domain channel changes slowly, the proposed scheme separates the channel into a dense vector and a sparse vector. By applying CS for the estimation of the sparse vector, the pilot overhead can be reduced when the channel is not sparse enough and conventional CS algorithms are not applicable. Numerical results verified that the proposed scheme can shorten training sequence significantly without explicit update of channel statistics.

Appendix
Proof of Proposition 1

Following the definitions, the MSE given the support can be decomposed as

$$\begin{aligned} \mathrm{MSE} &= \mathbb{E}\left[\|\mathbf{h}_i - \hat{\mathbf{h}}_i\|_2^2\right] \\ &= \mathbb{E}\left[\|\Psi_{\Omega_{i-1}}(\mathbf{s}_{i,d} - \hat{\mathbf{s}}_{i,d}) + \Psi_{\Omega_{i-1}^c}(\mathbf{s}_{i,s} - \hat{\mathbf{s}}_{i,s})\|_2^2\right] \\ &= \mathbb{E}\left[\|\mathbf{s}_{i,d} - \hat{\mathbf{s}}_{i,d}\|_2^2\right] + \mathbb{E}\left[\|\mathbf{s}_{i,s} - \hat{\mathbf{s}}_{i,s}\|_2^2\right]. \end{aligned} \tag{37}$$

The first term can be expressed as

$$\begin{aligned} \mathbb{E}\left[\|\mathbf{s}_{i,d} - \hat{\mathbf{s}}_{i,d}\|_2^2\right] &= \mathbb{E}\left[\|(\mathbf{s}_{i,d} - \tilde{\mathbf{s}}_{i,d})_{\Lambda_{i,d}}\|_2^2\right] \\ &= \frac{1}{\rho T}\mathbb{E}\left[\|(\tilde{\mathbf{X}}_d \mathbf{z}_{i,d}^*)_{\Lambda_{i,d}}\|_2^2\right] \\ &= \frac{k - k_s}{\rho T}, \end{aligned} \tag{38}$$

using (22).

For the second term, recall that $\hat{\mathbf{s}}_{i,s}$ is the output of the OMP algorithm. According to the result of [32], the support of $\mathbf{s}_{i,d}$ is exactly recovered by OMP under the assumption of RIP and sufficiently large SNR. Given the exact support recovery, the OMP reconstructs the sparse signal with the LS approach. As a result, the output of the OMP becomes

$$\hat{\mathbf{s}}_{i,s} = \frac{1}{\sqrt{\rho T}}(\mathbf{I}_{M-k})_{\Lambda_{i,s}}\left(\Phi_{\Lambda_{i,s}}^H \Phi_{\Lambda_{i,s}}\right)^{-1}\Phi_{\Lambda_{i,s}}^H \mathbf{y}_{i,s}^*. \tag{39}$$

With this, we can write the second term in (37) as

$$\begin{aligned} &\mathbb{E}\left[\|\mathbf{s}_{i,s} - \hat{\mathbf{s}}_{i,s}\|_2^2\right] \\ &= \frac{1}{\rho T}\mathbb{E}\left[\|(\mathbf{I}_{M-k})_{\Lambda_{i,s}}(\Phi_{\Lambda_{i,s}}^H \Phi_{\Lambda_{i,s}})^{-1}\Phi_{\Lambda_i}^H \mathbf{z}_{i,s}^*\|_2^2\right] \\ &= \frac{1}{\rho T}\mathrm{tr}\big((\Phi_{\Lambda_{i,s}}^H \Phi_{\Lambda_{i,s}})^{-1}\big). \end{aligned} \tag{40}$$

Combining (38) and (40), the equality of (27) can be obtained.

With a trace property, we have

$$\mathrm{tr}\big((\Phi_{\Lambda_i}^H \Phi_{\Lambda_i})^{-1}\big) = \sum_{j=1}^{k_s} \frac{1}{\lambda_j}, \tag{41}$$

where the λ_j's are the eigenvalues of $\Phi_{\Lambda_i}^H \Phi_{\Lambda_i}$. Since $\bar{\Phi}$ satisfies the RIP of order k_s+1, it is obvious that the eigenvalues are bounded as

$$\frac{T_p - k}{M - k}(1 - \delta_{k_s+1}) \leq \lambda_j \leq \frac{T_p - k}{M - k}(1 + \delta_{k_s+1}), \quad \text{for all } j, \tag{42}$$

resulting in

$$\frac{(M-k)k_s}{\rho_T(T_p - k)(1+\delta_{k_s+1})} \leq \mathbb{E}\left[\|\mathbf{s}_{i,s} - \hat{\mathbf{s}}_{i,s}\|_2^2\right] \leq \frac{(M-k)k_s}{\rho_T(T_p - k)(1-\delta_{k_s+1})}. \tag{43}$$

Using (38) and (43), we can obtain the bound of MSE.

REFERENCES

[1] F. Boccardi, R. W. Heath, Jr., A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

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

[3] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[4] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[5] C. K. Au-Yeung, D. J. Love, and S. Sanayei, "Trellis coded line packing: Large dimensional beamforming vector quantization and feedback transmission," *IEEE Trans. Wireless Commun.*, vol. 10, no. 6, pp. 1844–1853, Jun. 2011.

[6] J. Choi, Z. Chance, D. J. Love, and U. Madhow, "Noncoherent trellis coded quantization: A practical limited feedback technique for massive MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 12, pp. 5016–5029, Dec. 2013.

[7] J. Choi, D. J. Love, and T. Kim, "Trellis-extended codebooks and successive phase adjustment: A path from LTE-advanced to FDD massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 2007–2016, Apr. 2015.

[8] Y. Han, W. Shin, and J. Lee, "Projection-based differential feedback for FDD massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 1, pp. 202–212, Jan. 2017.

[9] 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.

[10] S. Noh, M. D. Zoltowski, Y. Sung, and D. J. Love, "Pilot beam pattern design for channel estimation in massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 787–801, Oct. 2014.

[11] J. Choi, D. J. Love, and P. Bidigare, "Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 802–814, Oct. 2014.

[12] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[13] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[14] E. J. Candés, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[15] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.

[16] C. R. Berger, Z. Wang, J. Huang, and S. Zhou, "Application of compressive sensing to sparse channel estimation," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 164–174, Nov. 2010.

[17] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, Jun. 2010.

[18] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[19] Y. Zhou, M. Herdin, A. M. Sayeed, and E. Bonek, "Experimental study of MIMO channel statistics and capacity via the virtual channel representation," Dept. ECE, Univ. Wisconsin-Madison, Madison, WI, USA, Tech. Rep., Feb. 2007.

[20] S. L. H. Nguyen and A. Ghrayeb, "Compressive sensing-based channel estimation for massive multiuser MIMO systems," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Shanghai, China, Apr. 2013, pp. 2890–2895.

[21] X. Rao and V. K. N. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3261–3271, Jun. 2014.

[22] P.-H. Kuo, H. T. Kung, and P.-A. Ting, "Compressive sensing based channel feedback protocols for spatially-correlated massive antenna arrays," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Paris, France, Apr. 2012, pp. 492–497.

[23] 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.

[24] 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.

[25] T. Kim and D. J. Love, "Virtual AoA and AoD estimation for sparse millimeter wave MIMO channels," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Stockholm, Sweden, Jun./Jul. 2015, pp. 146–150.

[26] Y. Han and J. Lee, "Two-stage compressed sensing for millimeter wave channel estimation," in *Proc. IEEE Int. Symp. Inf. Theory*, Barcelona, Spain, Jul. 2016, pp. 860–864.

[27] S. Wu, C. X. Wang, H. Haas, E. H. M. Aggoune, M. M. Alwakeel, and B. Ai, "A non-stationary wideband channel model for massive MIMO communication systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1434–1446, Mar. 2015.

[28] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[29] X. Rao and V. K. N. Lau, "Compressive sensing with prior support quality information and application to massive MIMO channel estimation with temporal correlation," *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4914–4924, Sep. 2015.

[30] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

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

[32] R. Wu, W. Huang, and D.-R. Chen, "The exact support recovery of sparse signals with noise via orthogonal matching pursuit," *IEEE Signal Process. Lett.*, vol. 20, no. 4, pp. 403–406, Apr. 2013.

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

[34] A. S. Bandeira, E. Dobriban, D. G. Mixon, and W. F. Sawin, "Certifying the restricted isometry property is hard," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3448–3450, Jun. 2013.

[35] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approx.*, vol. 28, no. 3, pp. 253–263, Dec. 2008.

[36] M. Elad, "Optimized projections for compressed sensing," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5695–5702, Dec. 2007.

[37] G. Li, Z. Zhu, D. Yang, L. Chang, and H. Bai, "On projection matrix optimization for compressive sensing systems," *IEEE Trans. Signal Process.*, vol. 61, no. 11, pp. 2887–2898, Jun. 2013.

[38] P. Kyritsi, D. C. Cox, R. A. Valenzuela, and P. W. Wolniansky, "Correlation analysis based on MIMO channel measurements in an indoor environment," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 713–720, Jun. 2003.

[39] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, Jul. 2005.

[40] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, Mar. 2006.

[41] N. Vaswani, "LS-CS-residual (LS-CS): Compressive sensing on least squares residual," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4108–4120, Aug. 2010.

[42] N. Vaswani and W. Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4595–4607, Sep. 2010.

[43] M. P. Friedlander, H. Mansour, R. Saab, and Ö. Yilmaz, "Recovering compressively sampled signals using partial support information," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1122–1134, Feb. 2012.

[44] J. Ziniel and P. Schniter, "Dynamic compressive sensing of time-varying signals via approximate message passing," *IEEE Trans. Signal Process.*, vol. 61, no. 21, pp. 5270–5284, Nov. 2013.

[45] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci. USA*, vol. 106, pp. 18914–18919, Nov. 2009.

[46] J.-C. Shen, J. Zhang, E. Alsusa, and K. B. Letaief, "Compressed CSI acquisition in FDD massive MIMO: How much training is needed?" *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 4145–4156, Jun. 2016.

[47] D. Ying, F. W. Vook, T. A. Thomas, D. J. Love, and A. Ghosh, "Kronecker product correlation model and limited feedback codebook design in a 3D channel model," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Sydney, NSW, Australia, Jun. 2014, pp. 5865–5870.

[48] Z. Marzi, D. Ramasamy, and U. Madhow, "Compressive channel estimation and tracking for large arrays in mm-wave picocells," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 514–527, Apr. 2016.

Yonghee Han received the B.S. and Ph.D. degrees in electrical engineering and computer science from Seoul National University, Seoul, South Korea, in 2011 and 2017, respectively. In 2015, he held a visiting position with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA. He is currently a Post-Doctoral Researcher with the Institute of New Media and Communications, Seoul National University.

His research interests are in the design of wireless communication systems, with emphasis on massive MIMO and millimeter wave communication systems. He received the Bronze Prize in 2016 Samsung HumanTech Paper Contest.

Jungwoo Lee received the B.S. degree in electronics engineering from Seoul National University, Seoul, South Korea, in 1988, and the M.S.E. and Ph.D. degrees in electrical engineering from Princeton University in 1990 and 1994, respectively. He was a Member of Technical Staff involved in multimedia signal processing with SRI (Sarnoff) from 1994 to 1999, where he was a Team Leader (PI) for an $18M NIST ATP Program. He has been with the Wireless Advanced Technology Lab, Lucent Technologies Bell Labs, since 1999, and involved in W-CDMA base station algorithm development as a Team Leader, for which he received two Bell Labs technical achievement awards. He is currently a Professor with the Department of Electrical and Computer Engineering, Seoul National University. He holds 21 U.S. patents. His research interests include wireless communications, information theory, distributed storage, and machine learning. He has also been the Track Chair of the IEEE ICC SPC from 2016 to 2017, and was a TPC/OC Member for ICC15, ITW15, VTC15s, ISIT09, PIMRC08, ICC05, and ISITA05. He received the Qualcomm Dr. Irwin Jacobs Award in 2014 for his contributions in wireless communications. He was an Associate Editor of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY from 2008 to 2011 and the *Journal of Communications and Networks* from 2012 to 2016. He is an Editor of the IEEE WIRELESS COMMUNICATIONS LETTERS. He has also been a Chief Editor of the *KICS Journal* and an Executive Editor of the *ICT Express* (Elsevier-KICS) since 2015.

David J. Love received the B.S. (Highest Honors), M.S.E., and Ph.D. degrees in electrical engineering from The University of Texas at Austin in 2000, 2002, and 2004, respectively. Since 2004, he has been with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA, where he is currently a Professor and recognized as a University Faculty Scholar. He has been very involved in commercialization of his research and holds 27 issued U.S. patents. He has served as an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS, an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and a Guest Editor for special issues of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS and the *EURASIP Journal on Wireless Communications and Networking*.

Dr. Love has been recognized as a Thomson Reuters Highly Cited Researcher. Along with his co-authors, he received best paper awards from the IEEE Communications Society (2016 IEEE Communications Society Stephen O. Rice Prize), the IEEE Signal Processing Society (2015 IEEE Signal Processing Society Best Paper Award), and the IEEE Vehicular Technology Society (2009 IEEE Transactions on Vehicular Technology Jack Neubauer Memorial Award). He has received multiple IEEE Global Communications Conference (Globecom) best paper awards.