Compressed CSI Acquisition in FDD Massive MIMO: How Much Training is Needed?

Juei-Chin Shen, *Member, IEEE*, Jun Zhang, *Senior Member, IEEE*, Emad Alsusa, *Senior Member, IEEE*, and Khaled B. Letaief, *Fellow, IEEE*

***Abstract*—Massive multiple-input–multiple-output (MIMO) is a promising technique for providing unprecedented spectral efficiency. However, it has been well recognized that the excessive training overhead required for obtaining the channel side information is a major handicap in frequency-division duplexing (FDD) massive MIMO. Several attempts have been made to reduce this training overhead by exploiting the sparsity structures of massive MIMO channels. So far, however, there has been little discussion about how to exploit the partial support information of these channels to achieve further overhead reductions. Such information, which is a set of indices of the significant elements of a channel vector, can be acquired in advance and hence is an important option to explore. In this paper, we examine the impact on the required training overhead when this information is applied within a weighted ℓ_1 minimization framework, and analytically show that a sharp estimate of the reduced overhead size can be successfully obtained. Furthermore, we examine how the accuracy of the partial support information impacts the achievable overhead reduction. Numerical results for a wide range of sparsity and partial support information reliability levels are presented to quantify our findings and main conclusions.**

***Index Terms*—Massive MIMO, channel estimation, pilot contamination, FDD, compressed sensing, weighted ℓ_1 minimization, partial support information, phase transition.**

I. Introduction

In massive MIMO, reaping the full benefit of excessive base station (BS) antennas requires perfect knowledge of the channel side information (CSI) at the BS [2]. Unfortunately, it has been indicated that the downlink pilot training for acquiring CSI consumes a significant portion of the radio resources in FDD massive MIMO [3]. A straightforward way to circumvent this difficulty is by employing the time-division duplexing (TDD) mode and exploiting the channel reciprocity via uplink training. In this way, the training overhead becomes proportional to the number of user equipments (UEs) instead of the BS antenna array size. Nonetheless, TDD massive MIMO encounters a serious problem as the number of UEs increases, due to the corresponding increase in the reuse of uplink pilots [4]. This reuse eventually causes inter-user interference and hence limits the system performance.

A recent study found that low-overhead pilot training is still feasible in the presence of dimension shrinkage of correlated massive MIMO channels [5]. It means that the effective dimension of a channel can be greatly less than its original dimension, which results in a lighter pilot training burden. In the case of a large uniform linear array (ULA) at the BS, it was shown that this effective channel dimension is governed by antenna spacing, angle of departure (AoD), and angle spread (AS) [5, Theorem 2]. Essentially, this dimension becomes smaller with the increased spatial correlation due to decreasing AS. The AS is normally equal to 5° in urban areas [6], which implies dimension shrinkage up to 90%. This outcome stems from the classical one-ring model [7] which is corroborated by experimental data [8]. A similar conclusion has been reached independently in [9], where a classical multipath model also exhibits the dimension shrinkage that occurs in massive MIMO.

The two basic assumptions, namely the far-field and plane wavefronts, underpinning the one-ring model could be violated due to the physical size of large antenna arrays [10]. As indicated by the channel measurements in [11], [12], the number of AoDs and the range of AS can potentially increase with the growing antenna size. This raises a concern that the effective channel dimension can be higher than expected. However, the measurement experiments were conducted in the environment with local scatterers around the BS [11], [12], which may be an enabling and indispensable factor that makes an unexpected increase in the AoDs or AS possible. Because this environment is not considered in the one-ring model[1], the aforementioned measurements are not sufficient to disprove the dimension-shrinkage result derived from this model. In addition, the far-field and plane wavefront assumptions are still reasonable as massive MIMO operates at higher radio frequencies such as the millimeter-wave bands.

Taking advantage of dimension shrinkage, several methods have shown their potentials in low-overhead pilot training.

Manuscript received July 5, 2015; revised January 2, 2016; accepted February 10, 2016. Date of publication February 26, 2016; date of current version June 7, 2016. This work was supported by the Hong Kong Research Grant Council under Grant 16211815. This work was presented in part at the IEEE International Conference on Communications [1]. The associate editor coordinating the review of this paper and approving it for publication was A. Wyglinski.

J.-C. Shen is with MediaTek Inc., Hsinchu 30078, Taiwan (e-mail: jc.shen@mediatek.com).

J. Zhang and K. B. Letaief are with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Kowloon, Hong Kong (e-mail: eejzhang@ust.hk; eekhaled@ust.hk).

K. B. Letaief is also with Hamad Bin Khalifa University, Qatar (e-mail: kletaief@hbku.edu.qa).

E. Alsusa is with the School of Electrical and Electronic Engineering, University of Manchester, Manchester M13 9PL, U.K. (e-mail: e.alsusa@manchester.ac.uk).

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/TWC.2016.2535310

[1]It means that we implicitly exclude massive MIMO channels in the environment with local scatterers around BSs from consideration. Nevertheless, this kind of channel is also an important category that requires further research.

A correlation based method provides a means of determining the optimal pilot sequences according to well-established criteria [13]. In [14], it has been demonstrated that the corresponding pilot length can be significantly smaller than the number of the BS antennas in correlated channels. However, certain drawbacks associated with the use of this method can outweigh its benefits when the dimensions of the channel covariance matrices that need to be acquired become large [15, Remark 2]. Another effective way that can be used is based on the discrete Fourier transform (DFT) pre-beamforming, where large ULAs are employed and only angular domain information is required [5]. Pre-beamforming at the BS, however, can potentially make the desired channel estimate contaminated by the interfering channel when using unreliable angular domain information.

In this paper, we focus on CSI acquisition of correlated massive MIMO channels where dimension shrinkage takes place. The aforementioned drawback of the pre-beamforming method is resolved by taking advantage of state-of-the-art compressed sensing techniques. In this regard, channel recovery is formulated as a weighted ℓ_1 minimization problem, where channel support knowledge[2] is utilized at the UE side. In particular, the weighting coefficients will depend on channel support information. More importantly, the uncertainty of channel support information will be taken into account in our problem formulation. By further conducting a convex geometry analysis, we are able to address a crucial question, namely, given the original channel dimensions, channel support information, and certain accuracy level of this information, *how much training is needed to guarantee meaningful channel estimates?* Our study shows that an upper bound on a geometric object remarkably provides sharp estimates of the required training overhead with a certain degree of probabilistic guarantee. We also find that if a certain level of support information accuracy is not achieved, incorporating support information into channel recovery does not result in any benefit.

A. Related Works

How much training is needed has long been a question of great interest in the FDD MIMO systems. The classical result shows that the optimal number of training symbols, for maximizing an achievable capacity bound, is equal to the number of BS antennas [16]. For optimizing the degrees of freedom, half of the channel coherence interval should be allocated to training [17]. When taking CSI feedback and beamforming into account, how to scale the training interval to the optimal length is provided in [18]. However, the assumption of channel entries being independent and identically distributed (i.i.d.) Gaussian makes these results less generalizable to correlated MIMO channels as in typical massive MIMO systems.

Much of the recent investigation into training for correlated channels has centered on training overhead lessening by exploiting prior channel knowledge. In [19], temporal and spatial channel correlations are utilized to facilitate open/closed loop training, thereby leading to reduced downlink training overhead. Compressed sensing provides a framework for efficient CSI acquisition utilizing prior knowledge of channel sparsity structures [20]. To adaptively exploit these sparsity structures according to their quality, a modified subspace pursuit (M-SP) algorithm has recently been developed [21]. The analytic performance of this algorithm has also been derived via a restricted isometry property (RIP) approach. It was shown in [22], [23] that if massive MIMO is built upon an orthogonal frequency division multiplexing (OFDM) system, the common support of channel impulse responses (CIRs) is another source of utilizable sparsity structures. Though a variety of compressed sensing based methods have been developed, few of them attemp to quantify the required training as a function of the available channel knowledge.

[2]The support of a channel vector contains the indexes of significant channel elements in the angular domain.

B. Contributions

The major contributions of this paper are summarized as follows:

1) We have developed a weighted ℓ_1 minimization framework that can incorporate statistical channel knowledge in the form of partial support information, thus harnessing sparsity structures in the angular domain to achieve low-overhead training. In particular, this approach circumvents the drawback of the existing pre-beamforming method as the statistical knowledge is utilized at the UE side. One interesting finding is that how significantly the training overhead can be reduced depends on the accuracy of the prior knowledge. This finding, while preliminary, suggests that a method has to be developed to deal with the situation where statistical channel knowledge becomes unreliable.
2) We are the first to establish precise bounds on the number of linear channel measurements that are necessary for sparse massive MIMO channel recovery using weighted ℓ_1 minimization. The bounds, specified by Theorem 5 and Proposition 8, are not restricted to the asymptotic region, where the channel dimension has to be sufficiently large, and take the uncertainty of the statistical channel knowledge into account. This allows us to provide sharp estimates of the required training overhead, given a certain accuracy level of the statistical knowledge and a modest channel dimension size. On the other hand, offering these estimates has important implications for determining what degree of accuracy of the prior knowledge is needed and evaluating the cost of the corresponding training.

C. Organization and Notations

The remainder of this paper is organized as follows. In Section II, we specify the system model including the sparse MIMO channels and their partial support information. Section III addresses the issue of sparse channel reconstruction using weighted ℓ_1 minimization and interference cancellation in the compressed domain. Quantifying the required training overhead is conducted in Section IV. Numerical results are

presented in Section V, which is followed by the concluding remarks in Section VI.

Notations: $\mathbb{R}$: real number, $\mathbb{C}$: complex number, $\mathbb{Z}$: integers, $\Re\{\cdot\}$: real part, $\Im\{\cdot\}$: imaginary part, $\|\cdot\|_p$: p-norm, $(\cdot)'$: transpose, $(\cdot)^H$: Hermitian transpose, $\mathbf{I}_N$: $N \times N$ identity matrix, $\mathcal{N}(\cdot,\cdot)$: normal distribution, $\mathcal{CN}(\cdot,\cdot)$: complex normal distribution, $\mathbb{E}[\cdot]$: expectation, $\mathbf{0}$: zero vector, supp($\mathbf{x}$): the set of indices i such that the entry x_i of the vector $\mathbf{x}$ is non-zero, card $(\cdot)$: cardinality.

II. System Model

Consider a massive MIMO system consisting of a BS and K single-antenna user equipments. The BS is equipped with an M-element uniform linear array (ULA), where M is supposed to be greatly larger than the number of served UEs. The system is assumed to operate in the FDD mode and the channels between the BS and the UEs are flat block-fading. During the downlink training phase, the received signals at a typical UE can be represented as

$$\mathbf{y} = \mathbf{A}\mathbf{h} + \mathbf{e}, \tag{1}$$

where $\mathbf{A} \in \mathbb{C}^{N\times M}$ denotes the training matrix, $\mathbf{h} \in \mathbb{C}^{M\times 1}$ is the channel vector, and $\mathbf{e} \in \mathbb{C}^{N\times 1}$ is the additive error due to noise or interference. The conventional least-square (LS) estimator requires the number N of channel measurements to scale linearly with the number of BS antennas, which is prohibitively large.

A. Sparse Massive MIMO Channels

The high multiplexing gain achieved in massive MIMO results from its fine spatial resolution [2]. Hence, it is reasonable to have BS antennas half-carrier-wavelength spaced to provide high resolution in azimuth. The way to realize such antenna deployment has been envisioned in [24]. Based on this array manifold, the quasi-static channel $\mathbf{h}$ can be canonically decomposed as

$$\mathbf{h} = \mathbf{U}\mathbf{h}^{\mathrm{a}}, \tag{2}$$

where $\mathbf{h}^{\mathrm{a}}$ is the channel representation in the angular domain and $\mathbf{U}$ is an $M \times M$ discrete Fourier transform (DFT) matrix [25]. The nth column of $\mathbf{U}$ given by

$$\mathbf{u}_n \triangleq \frac{1}{\sqrt{M}}\left[\,1 \;\cdots\; e^{-j\frac{2\pi(m-1)n}{M}} \;\cdots\; e^{-j\frac{2\pi(M-1)n}{M}}\,\right]' \tag{3}$$

stands for an angular basis vector. The channel $\mathbf{h}$ due to the effect of clusters of scatterers can be written as [26]

$$\mathbf{h} = \sum_{p=1}^{P} g_p \mathbf{s}(\theta_p), \tag{4}$$

where P is the number of scatterers, g_p denotes the complex path gain, $\theta_p \triangleq \theta + \delta_p$ represents the AoDs, θ is the mean AoD, and δ_p is the random angular deviation. $\mathbf{s}(\theta) \triangleq \frac{1}{\sqrt{M}}[\,1 \;\; e^{-j\pi\cos(\theta)} \;\cdots\; e^{-j\pi(M-1)\cos(\theta)}\,]'$ is the steering vector in the direction of θ. Based on this channel representation, it has been shown that dimension shrinkage occurs due to the limited range of AoDs [9]. How this dimension shrinkage implies that the support of $\mathbf{h}^{\mathrm{a}}$ being small is provided below.

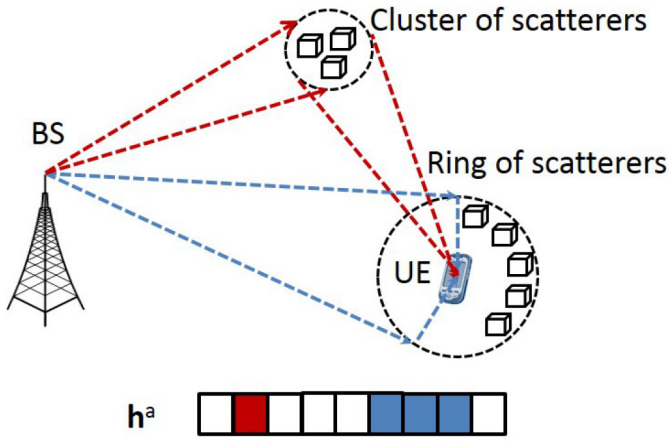


Fig. 1. Illustration of the sparse channel in the angular domain.

Recasting (4) in the canonical form gives

$$\mathbf{h}^{\mathrm{a}} = \mathbf{U}^H\mathbf{S}\mathbf{r}, \tag{5}$$

where $\mathbf{S} = [\mathbf{s}(\theta_1), \cdots, \mathbf{s}(\theta_P)]$ and $\mathbf{r} = [g_1, \cdots, g_P]'$. The inner product

$$\mathbf{u}_n^H\mathbf{s}(\theta) = \frac{\sin[\pi M\Omega(n,\theta)]}{M\sin[\pi\Omega(n,\theta)]}e^{j\pi(M-1)\Omega(n,\theta)} \tag{6}$$

with $\Omega(n,\theta) = (n/M - \cos\theta/2)$ shows a significant magnitude if $|\Omega(n,\theta)| < 1/M$. In other words, the path of direction θ is resolved by the nth angular basis vector if $(n-1)/M < \cos\theta/2 < (n+1)/M$. Specifically, the nth resolvable channel gain h_n^{a} results from an aggregate of multiple physical paths with AoDs θ_p's which satisfy the condition $|\Omega(n,\theta_p)| < 1/M$. Associating the AoDs with their corresponding resolvable channel gains provides the support information of $\mathbf{h}^{\mathrm{a}}$, which is denoted by $T = \mathrm{supp}(\mathbf{h}^{\mathrm{a}})$, where $T \subseteq \{1,\ldots,M\}$ and $\mathrm{card}(T) = s$. The angular deviation δ_p can be modeled as a Gaussian variate with distribution $\mathcal{N}(0, \sigma_\delta^2)$ or a uniform variate which lies within $[-\sqrt{3}\sigma_\delta, \sqrt{3}\sigma_\delta]$ [27], [28], where σ_δ is referred to as the AS deviation. As shown in Fig. 1, when the BS antenna array is mounted higher than the nearby scatterers, multipaths are mainly due to the far-field scatterers or those in the vicinity of the target UE. In this case, we expect the cardinality of the angular channel support to be relatively small in comparison to the actual channel dimension.

B. Partial Support Information

To obtain partial support information, we can either directly estimate AoDs or indirectly evaluate the mean AoD and the AS deviation. Most of the existing methods for obtaining the aforementioned spatial information require knowledge of the downlink channel covariance matrices. To meet this prerequisite, one has to consume lots of radio resources in terms of

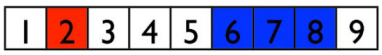


Fig. 2. An illustrative example of partial support information, where $T = \{2, 6, 7, 8\}$, $\widehat{T} = \{2, 3, 6, 7, 8, 9\}$, and $\text{card}\left(T \cap \widehat{T}\right) = \lfloor \alpha \times \hat{s} \rfloor = \frac{2}{3} \times 6 = 4$.

feeding back downlink channel measurements from UEs to BSs. Due to spatial reciprocity in FDD systems [29], the spatial behaviors are closely related at the downlink and uplink carrier frequencies. In addition, it has been shown in [30] that the downlink channel covariance matrix can be determined by the uplink channel covariance matrix via certain calibration processing given uplink and downlink carrier frequencies, and the BS antenna geometry. That is, the desired spatial information can be reasonably retrieved during uplink reception.

The use of subspace-based techniques, such as the multiple signal classification (MUSIC) and the estimation of signal parameters via rotational invariance techniques (ESPRIT) algorithms [31], is well-established in high-resolution AoD estimation. Integrated with the spatial signature estimation algorithm, subspace-based techniques are capable of resolving $2M/3$ AoDs for each channel to a UE [32]. As a result, if the number of BS antennas is sufficiently large, i.e., $M \geq 3P/2$, it is possible to achieve accurate AoD estimation, which was regarded as unattainable in conventional MIMO systems [30]. Recent developments in the field of compressed sensing have led to a renewed interest in AoD estimation by finding the sparsest representation in an overcomplete angular basis [33]. It is indicated in [34] that this approach outperforms the eigenstructure-based ones in terms of being more adaptive in various circumstances such as the operating signal-to-noise ratio (SNR) being low or spatial samples being limited.

In [27], a decoupled estimation of the mean AoD and the AS deviation is proposed based on the covariance matrix matching approach. The dimension of the required sample covariance matrix can be greatly less than M for the purpose of estimating the mean AoD and the AS. Hence, the required sample size for estimating the channel covariance matrix will be greatly lessened. In the case of angular deviations being uniformly distributed, the estimates $\hat{\theta}$ and $\hat{\sigma}_\delta$ will determine a range $[\hat{\theta} - \sqrt{3}\hat{\sigma}_\delta, \hat{\theta} + \sqrt{3}\hat{\sigma}_\delta]$ within which the AoDs θ_p are likely to lie. More advanced techniques for jointly estimating the mean AoDs and the ASs of several spatially distributed sources have been presented in [35], which allow simultaneous spatial information acquisition for multiple UEs.

Associating the AoD or AoD range estimates with their corresponding resolvable paths provides partial support information of $\mathbf{h}^{\text{a}}$, which is denoted by $\widehat{T}$, where $\widehat{T} \subseteq \{1, \ldots, M\}$ and $\text{card}(\widehat{T}) = \hat{s}$. The intersection $T \cap \widehat{T}$ of the desired and estimated channel supports has the size $\text{card}(T \cap \widehat{T}) = \lfloor \alpha \hat{s} \rfloor$, where $0 \leq \lfloor \alpha \hat{s} \rfloor \leq s$ and $0 \leq \hat{s} - \lfloor \alpha \hat{s} \rfloor \leq M - s$. An illustrative example is provided in Fig. 2. The parameter $\alpha \in [0, 1]$ indicates the accuracy or quality of the partial support information $\widehat{T}$, i.e., $\alpha = 1$ implies that all the indices in $\widehat{T}$ are accurate, whereas for $\alpha = 0$ the partial support information is completely irrelevant to the actual channel support.

III. Compressed CSI Acquisition Using Partial Support Information

It has been recognized that the amount of training overhead, which is proportional to the BS antenna size, makes massive MIMO difficult to implement in the FDD mode. However, when the channel $\mathbf{h}^{\text{a}}$ exhibits a sparsity structure ($s \ll M$), low-overhead training becomes feasible by exploiting compressed sensing (CS) techniques [20]. With further utilization of partial support information, it will be shown that an additional reduction in training overhead can be achieved.

A. Weighted ℓ_1 Minimization with Partial Support Information

In this section, partial support information will be applied within the framework of weighted ℓ_1 minimization, which has demonstrated its great potential of lessening the required training overhead [36]–[38]. Given linear channel observations

$$\mathbf{y} = \mathbf{AUh}^{\text{a}} + \mathbf{e}, \tag{7}$$

the channel recovery can be formulated as a weighted ℓ_1 minimization problem given by

$$(\mathcal{P}1) \quad \begin{aligned} &\min_{\hat{\mathbf{h}}^{\text{a}} \in \mathbb{C}^M} \left\|\hat{\mathbf{h}}^{\text{a}}\right\|_{1,\mathbf{w}} \text{ subject to } \left\|\mathbf{AU}\hat{\mathbf{h}}^{\text{a}} - \mathbf{y}\right\|_2 \leq \epsilon, \\ &\text{with} \qquad\qquad w_i = \begin{cases} 1, & i \notin \widehat{T}, \\ 0, & i \in \widehat{T}, \end{cases} \end{aligned} \tag{8}$$

where $\mathbf{A} \in \mathbb{C}^{N \times M}$ is a Gaussian random matrix of random entries with distribution $\mathcal{CN}(0, 1/N)$, the error $\mathbf{e}$ is assumed to be upper bounded, i.e., $\|\mathbf{e}\|_2 \leq \epsilon$, and $\left\|\hat{\mathbf{h}}^{\text{a}}\right\|_{1,\mathbf{w}} \triangleq \sum_{i=1}^{M} w_i \left|\hat{h}_i^{\text{a}}\right|$. The main idea behind this formulation is that the entries that are expected to be zero should be weighted more heavily than others in the objective function. As in [39], the measurement matrix $\mathbf{A}$ can be generated offline and made accessible to the BSs and the UEs. Meanwhile, the support estimate $\widehat{T}$ is also made available to the corresponding UE to enable it to recover the channel and feed back sparse channel information $\hat{\mathbf{h}}^{\text{a}}$ to its serving BS.

It is intriguing to know the required number of measurements for which the mismatch $\left\|\hat{\mathbf{h}}^{\text{a}} - \mathbf{h}^{\text{a}}\right\|_2$ is negligible. Moreover, this number will provide an insight into the cost of different combinations of the parameters $\{M, s, \hat{s}, \alpha, \epsilon\}$. In Section IV, this number will be quantified by utilizing a convex geometry approach. This kind of quantification in compressed sensing is referred to *as phase transition analysis*.

B. Compressed Domain Interference Cancellation

If there is another BS performing the downlink training with the same training matrix, the additive error can have the form $\mathbf{e} = \mathbf{AUh}_I^{\text{a}}$, where $\mathbf{h}_I^{\text{a}}$ represents the angular interfering channel with support T_I. In this case, it is possible to remove the effect of interference via certain post-processing when T_I becomes available at the target UE[3]. For instance, we may replace $\widehat{T}$ in

[3]It is more reasonable to assume that only partial support information of the interfering channel is available. However, for the sake of simplicity, we assume that T_I is obtainable.

($\mathcal{P}1$) with $T_A \triangleq \widehat{T} \cup T_I$ to obtain an estimate of the composite channel $\left(\mathbf{h}^{\mathrm{a}} + \mathbf{h}^{\mathrm{a}}_I\right)$, and then remove the interference part from this estimate. This method, however, suffers from a potential increase in training overhead due to recovering the composite channel of a higher sparsity level.

Instead of reconstructing the composite channel first, it is desirable to mitigate interference in the compressed domain before recovering the target channel. In [40], a subspace-based method is proposed for compressed domain interference cancellation by exploiting the T_I information. Based on this, a modified method which takes advantage of prior knowledge of both $\widehat{T}$ and T_I is presented below. This prior knowledge first enables us to construct a projection matrix

$$\mathbf{B}^{\dagger}_{\widehat{T}} = \mathbf{B}_{\widehat{T}} \left(\mathbf{B}^{H}_{\widehat{T}} \mathbf{B}_{\widehat{T}}\right)^{-1} \mathbf{B}^{H}_{\widehat{T}} \tag{9}$$

where $\mathbf{B}_{\widehat{T}}$ is the $N \times \hat{s}$ matrix composed of the columns of $\mathbf{B} \triangleq \mathbf{AU}$ corresponding to the indices in $\widehat{T}$. Then, a basis of the space $C(\mathbf{B}_{T_I}) \setminus C(\mathbf{B}_{\widehat{T}})$ is given by $\mathbf{B}_{T_I \setminus \widehat{T}} = \mathbf{B}_{T_I} - \mathbf{B}^{\dagger}_{\widehat{T}} \mathbf{B}_{T_I}$. Using this, we can obtain the desired projection matrix

$$\mathbf{C} = \mathbf{I} - \mathbf{B}_{T_I \setminus \widehat{T}} \left(\mathbf{B}^{H}_{T_I \setminus \widehat{T}} \mathbf{B}_{T_I \setminus \widehat{T}}\right)^{-1} \mathbf{B}^{H}_{T_I \setminus \widehat{T}} \tag{10}$$

onto the space $C(\mathbf{B}) \setminus \left(C(\mathbf{B}_{T_I}) \setminus C(\mathbf{B}_{\widehat{T}})\right)$. Applying this projection matrix to (7) gives

$$\begin{aligned} \mathbf{y}_{\mathrm{C}} &= \mathbf{C}\mathbf{y} \\ &= \mathbf{CAU}\left(\mathbf{h}^{\mathrm{a}}_1 + \mathbf{h}^{\mathrm{a}}_I\right) + \mathbf{C}\mathbf{z}. \end{aligned} \tag{11}$$

In the end, the desired channel can be recovered by solving the problem ($\mathcal{P}1_{\mathrm{IC}}$) that resembles ($\mathcal{P}1$) with the constraint replaced by $\left\|\mathbf{CAU}\hat{\mathbf{h}}^{\mathrm{a}} - \mathbf{y}_{\mathrm{C}}\right\|_2 \leq \epsilon$.

If $\widehat{T} \supset T$, then

$$\begin{aligned} &\mathbf{CAU}\mathbf{h}^{\mathrm{a}} \\ &= \mathbf{C}\mathbf{B}_T \mathbf{h}^{\mathrm{a}}_T, \\ &= \left[\mathbf{I} - \mathbf{B}_{T_I \setminus \widehat{T}} \left(\mathbf{B}^{H}_{T_I \setminus \widehat{T}} \mathbf{B}_{T_I \setminus \widehat{T}}\right)^{-1} \mathbf{B}^{H}_{T_I \setminus \widehat{T}}\right] \mathbf{B}_{T_1} \mathbf{h}^{\mathrm{a}}_{T_1}, \\ &= \mathbf{AU}\mathbf{h}^{\mathrm{a}}, \end{aligned} \tag{12}$$

where $\mathbf{h}^{\mathrm{a}}_T$ denotes the $1 \times s$ vector obtained by retaining the entries of $\mathbf{h}^{\mathrm{a}}$ corresponding to the indices in T. It means that $\mathbf{y}_{\mathrm{C}}$ preserves adequate information about $\mathbf{h}^{\mathrm{a}}$ after the subspace projection if the most part of T is contained in the partial information $\widehat{T}$. Meanwhile, projecting the interference $\mathbf{AU}\mathbf{h}^{\mathrm{a}}_I$ onto the subspace $C(\mathbf{B}) \setminus \left(C(\mathbf{B}_{T_I}) \setminus C(\mathbf{B}_{\widehat{T}})\right)$ leads to $||\mathbf{CAU}\mathbf{h}^{\mathrm{a}}_I|| \leq ||\mathbf{AU}\mathbf{h}^{\mathrm{a}}_I||$ which implies that partial interference suppression can be achieved in the compressed domain. This suppression becomes exceptionally prominent, i.e.,

$$\mathbf{CAU}\mathbf{h}^{\mathrm{a}}_I = 0, \tag{13}$$

when the column space $C(\mathbf{B}_{T_I})$ is orthogonal to $C(\mathbf{B}_{\widehat{T}})$.

IV. Phase Transition Analysis

In this section, we analytically examine phase transitions of the weighted ℓ_1 minimization problem, which provides definite boundaries of the sample size versus sparsity level tradeoff curves. In other words, for a given sparsity level, we will quantify the corresponding sample size threshold above which a certain level of the channel recovery performance will be guaranteed.

It can be observed that the product $\mathbf{AU}$ has the same structure as the matrix $\mathbf{A}$ in terms of having independently Gaussian distributed entries. Hence, the matrix $\mathbf{U}$ in ($\mathcal{P}1$) can be ignored without loss of generality, and $\mathbf{h}^{\mathrm{a}}$ can be regarded as a sparse vector with respect to the canonical basis instead of the angular basis. Throughout this section, the existence of $\mathbf{U}$ will be intentionally omitted. With this in mind, the real-valued counterpart of the complex-domain problem ($\mathcal{P}1$) can be written as

$$(\mathcal{P}2) \quad \min_{\tilde{\mathbf{h}}^{\mathrm{a}} \in \mathbb{R}^{2M}} \left\|\tilde{\mathbf{h}}^{\mathrm{a}}\right\|_{2,1,\mathbf{w}} \text{ subject to } \left\|\widetilde{\mathbf{A}}\tilde{\mathbf{h}}^{\mathrm{a}} - \tilde{\mathbf{y}}\right\|_2 \leq \epsilon, \tag{14}$$

where the linear measurements in the real domain $\tilde{\mathbf{y}}$ is given by

$$\begin{aligned} \tilde{\mathbf{y}} &= \widetilde{\mathbf{A}}\tilde{\mathbf{h}}^{\mathrm{a}} + \tilde{\mathbf{e}}, \\ &= \begin{bmatrix} \Re\{\mathbf{A}\} & -\Im\{\mathbf{A}\} \\ \Im\{\mathbf{A}\} & \Re\{\mathbf{A}\} \end{bmatrix} \begin{bmatrix} \Re\{\mathbf{h}^{\mathrm{a}}\} \\ \Im\{\mathbf{h}^{\mathrm{a}}\} \end{bmatrix} + \begin{bmatrix} \Re\{\mathbf{e}\} \\ \Im\{\mathbf{e}\} \end{bmatrix}, \end{aligned} \tag{15}$$

and the weighted $\ell_{2,1}$ norm of $\tilde{\mathbf{h}}^{\mathrm{a}}$ is defined as

$$\left\|\tilde{\mathbf{h}}^{\mathrm{a}}\right\|_{2,1,\mathbf{w}} \triangleq \sum_{i=1}^{M} w_i \sqrt{\tilde{h}_i^2 + \tilde{h}_{i+M}^2}. \tag{16}$$

As shown in (15), the entries of the measurement matrix $\widetilde{\mathbf{A}}$ are constrained by a structure, instead of being generally independent of each other. In [41], the problem ($\mathcal{P}2$) has been indicated to be closely related to the block-sparse compressed sensing (BSCS) problem with block size 2, which is given by

$$(\mathcal{P}3) \quad \min_{\tilde{\mathbf{x}} \in \mathbb{R}^{2M}} \|\tilde{\mathbf{x}}\|_{2,1,\mathbf{w}} \text{ subject to } \left\|\overline{\mathbf{A}}\tilde{\mathbf{x}} - \tilde{\mathbf{y}}\right\|_2 \leq \epsilon, \tag{17}$$

where $\overline{\mathbf{A}} \in \mathbb{R}^{2N \times 2M}$ is a matrix of independent random entries with distribution $\mathcal{N}(0, 1/2N)$. Despite the difference between $\widetilde{\mathbf{A}}$ and $\overline{\mathbf{A}}$, it has been demonstrated that the phase transitions of ($\mathcal{P}2$) coincide with those of ($\mathcal{P}3$) when ignoring the weighting vector $\mathbf{w}$ and the error $\mathbf{e}$ [41]. Therefore, we argue that the phase transitions of ($\mathcal{P}2$) can be acquired via ($\mathcal{P}3$). The following analysis will focus on characterizing the phase transitions of ($\mathcal{P}3$) when partial support information is available. Then, it will be verified in Sec. V that the obtained phase transition curves accurately depict the transition behavior of the original problem ($\mathcal{P}2$), or equivalently ($\mathcal{P}1$).

Previous studies have based their phase transition analysis on the combinatorial geometry framework [42]. Recently, a simpler and more effective convex geometry approach have been developed and introduced to analyze the phase transition phenomenon. This approach has an attractive feature in terms of offering sharp estimates of the required number of measurements for robust recovery of sparse signals [43]. Before applying this approach, we define some technical terms as below which will facilitate the investigation into the intrinsic convex geometry of the BSCS problem.

Definition 1: A set $\mathcal{C}$ is said to be a *cone* if $\nu z \in \mathcal{C}$ for any $z \in \mathcal{C}$ and $\nu \geq 0$.

Definition 2: The *tangent cone* at some nonzero $\mathbf{x} \in \mathbb{R}^{2M}$ with respect to the scaled unit ball $\mathcal{B}_p(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^{2M} |\ \|\mathbf{y}\|_p \leq \|\mathbf{x}\|_p\}$ is given by

$$\mathcal{T}_p(\mathbf{x}) = \text{cone}\left\{\mathbf{z} - \mathbf{x} |\ \|\mathbf{z}\|_p \leq \|\mathbf{x}\|_p\right\}. \tag{18}$$

Definition 3: The set

$$\mathsf{N}_p(\mathbf{x}) = \left\{\mathbf{v} |\ \langle \mathbf{v},\ \mathbf{z} - \mathbf{x}\rangle \leq 0\, \forall \mathbf{z} \text{ such that } \|\mathbf{z}\|_p \leq \|\mathbf{x}\|_p\right\} \tag{19}$$

is called the *normal cone* at $\mathbf{x}$ with respect to $\mathcal{B}_p(\mathbf{x})$.

Definition 4: With respect to a set $\Psi \subset \mathbb{R}^M$, the *Gaussian width* is defined as

$$\omega(\Psi) = \mathbb{E}_{\mathbf{g}}\left[\sup_{\mathbf{z}\in\Psi} \mathbf{g}^T\mathbf{z}\right], \tag{20}$$

where $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

To simplify the illustration of how the defined geometric objects are exploited, let us first assume that $\mathbf{e} = \mathbf{0}$ and $\epsilon = 0$ in ($\mathcal{P}$3). Following this, we have $\tilde{\mathbf{x}} = \tilde{\mathbf{x}}^* \triangleq \tilde{\mathbf{h}}^{\text{a}}$ if and only if $\text{null}\left(\overline{\mathbf{A}}\right) \cap \mathcal{T}_{2,1,\mathbf{w}}(\tilde{\mathbf{x}}^*) = \{\mathbf{0}\}$ [43]. Gordon's escape through the mesh theorem provided a means of probabilistically characterizing the event that the random null space of $\overline{\mathbf{A}}$ misses the tangent cone $\mathcal{T}_{2,1,\mathbf{w}}(\tilde{\mathbf{x}}^*)$ [44]. The likelihood of this event is regulated by a result in Gordon's theory, i.e.,

$$\mathbb{E}\left[\min_{\mathbf{x}\in\Phi} \left\|\overline{\mathbf{A}}\mathbf{x}\right\|_2\right] \geq \lambda_{2N} - \omega(\Phi), \tag{21}$$

where $\Phi = \mathcal{T}_{2,1,\mathbf{w}}(\tilde{\mathbf{x}}^*) \cap \mathbb{S}^{2M}$, $\mathbb{S}^{2M} \subset \mathbb{R}^{2M}$ denotes the unit $2M$-sphere, and the expected length of a $2N$-dimensional Gaussian random vector is given by $\lambda_{2N} = \sqrt{2}\Gamma\left(\frac{2N+1}{2}\right)\Gamma\left(\frac{2N}{2}\right)$. This result, together with a concentration of the measure principle, leads to the following recovery guarantees of the BSCS problem with partial support information [43, Corollary 3.3].

Theorem 5: 1) Let the error $\tilde{\mathbf{e}}$ be zero and $\epsilon = 0$. Then, $\tilde{\mathbf{x}}^*$ is the unique optimum of the problem ($\mathcal{P}$3) with probability at least

$$1 - \exp(-\frac{1}{2}[\lambda_{2N} - \omega(\Phi)]^2) \tag{22}$$

given

$$2N \geq \omega(\Phi)^2 + 1. \tag{23}$$

2) If $\tilde{\mathbf{x}}^*$ represents the optimal solution of ($\mathcal{P}$3), then $\|\tilde{\mathbf{x}}^* - \tilde{\mathbf{x}}\|_2 \leq \frac{2\epsilon}{\nu}$ with probability at least

$$1 - \exp(-\frac{1}{2}[\lambda_{2N} - \omega(\Phi) - \sqrt{2N}\nu]^2) \tag{24}$$

given

$$2N \geq \frac{\omega(\Phi)^2 + \frac{3}{2}}{(1-\nu)^2}. \tag{25}$$

The above theorem rigorously establishes the existence of decisive thresholds for sparse channel recovery with certain probabilistic guarantee. Apparently, these thresholds beyond which the measurement sizes should be increased are determined by the Gaussian width. To state explicitly what probabilistic performance is assured, we provide a more interpretable result in the following corollary.

Corollary 6: The results of Theorem 5 hold with probability at least $\eta \in (0, 1)$ if

1)

$$2N \geq \left(\omega(\Phi) + \beta_\eta\right)^2 + 1, \tag{26}$$

2)

$$2N \geq \frac{\left(\omega(\Phi) + \beta_\eta\right)^2 + \frac{3}{2}}{(1-\nu)^2}, \tag{27}$$

where $\beta_\eta \triangleq \sqrt{-2\ln(1-\eta)}$.

Proof: Please refer to Appendix A. ■

The above corollary indicates that the required number of linear measurements depends on the desired probabilistic performance. A better probabilistic performance corresponds to a higher number of linear measurements. Whether in the above theorem or corollary, the Gaussian width $\omega(\Phi)$ still needs to be evaluated so that the sharp bound on the required number of measurements can be obtained. In the evaluation of $\omega(\Phi)$, the normal cone $\mathsf{N}_{2,1,\mathbf{w}}(\tilde{\mathbf{x}}^*)$ to be used is provided in the subsequent lemma.

Lemma 7: The normal cone $\mathsf{N}_{2,1,\mathbf{w}}(\tilde{\mathbf{x}}^*)$ at $\tilde{\mathbf{x}}^*$ with respect to $\mathcal{B}_{2,1,\mathbf{w}}(\tilde{\mathbf{x}}^*)$ can be expressed as

$$\begin{aligned}\mathsf{N}_{2,1,\mathbf{w}}\left(\tilde{\mathbf{x}}^*\right) = \Bigg\{&\mathbf{v} \in \mathbb{R}^{2M} |\ v_i = v_{i+M} = 0 \text{ for } i \in \widehat{T},\\ &v_i = t\frac{\tilde{x}_i^*}{|x_i^*|},\ v_{i+M} = t\frac{\tilde{x}_{i+M}^*}{|x_i^*|},\ \text{for } i \in T \setminus \widehat{T},\\ &\sqrt{v_i^2 + v_{i+M}^2} \leq t \text{ for } i \in T^c \setminus \widehat{T},\ \text{for } t \geq 0\Bigg\}.\end{aligned} \tag{28}$$

Proof: Please refer to Appendix B. ■

With the aid of this lemma, we are in a position to provide a bound on $\omega(\Phi)$.

Proposition 8: The upper bound of the square of Gaussian width $\omega(\Phi)$ in Theorem 5 is given by

$$\begin{aligned}\omega(\Phi)^2 \leq \inf_{t\geq 0} \Bigg\{&2\hat{s} + (s - \alpha\hat{s})\left(2 + t^2\right)\\ &+ [M - \hat{s} - (s - \alpha\hat{s})]\int_t^\infty (r-t)^2 r \exp\left(-\frac{r^2}{2}\right) dr\Bigg\}.\end{aligned} \tag{29}$$

The unique optimum t_{opt} which minimizes the right-hand side of (29) is the solution of

$$\int_t^\infty \left(\frac{r}{t} - 1\right) r \exp\left(-\frac{r^2}{2}\right) \text{d}r = \frac{s - \alpha\hat{s}}{M - \hat{s} - (s - \alpha\hat{s})}. \tag{30}$$

Proof: Please refer to Appendix C. ■

The upper bound provided in this proposition is numerically obtainable, and can be utilized in Theorem 5 and Corollary 6 to acquire the bounds on the measurement numbers. Although the phase transitions are analyzed for the BSCS problem, it will be demonstrated later that they are also applicable to the weighted ℓ_1 minimization problem in the complex domain. An existing work [45] on the phase transition of certain complex CS problem only provides asymptotic results which hold when M is sufficiently large. In contrast, the results presented here are precise and non-asymptotic.

Next, we argue how the phase transitions of $(\mathcal{P}1)$ will be affected by compressed-domain interference cancellation. Under the condition $\widehat{T} \supset T$, the fact $\mathbf{CAUh}^{\mathrm{a}} = \mathbf{AUh}^{\mathrm{a}}$ indicates that performing interference cancellation has no adverse effects in terms of diminishing the desired channel information. Especially, the guarantee $\|\mathbf{Ce}\| \leq \|\mathbf{e}\|$ ensures that the phase transitions of $(\mathcal{P}1_{\mathrm{IC}})$ will outperform those of the original problem. On the other hand, when $\widehat{T} \cap T \neq T$, projecting $\mathbf{y}$ onto the subspace can cause information loss. This is because any signal located within the subspace $C(\mathbf{B}_{T_I}) \setminus C(\mathbf{B}_{\widehat{T}})$, whose intersection with $C(\mathbf{B}_T)$ may be no longer empty, will be canceled. Though the discussion here is not general, it highlights that the impact of the compressed-domain processing on phase transitions can be made marginal given proper partial support information of the desired channel.

V. Numerical Results

In this section, numerical results are presented to illustrate the potential overhead reduction of using weighted ℓ_1 minimization with partial support information. Moreover, the accuracy of analytical bounds on the number of the required channel measurements is verified via comparison with empirical phase transition curves.

A. Overhead Reduction by Weighted ℓ_1 Minimization

In Fig. 3, the impact of the parameter α on the normalized mean-square error (NMSE) performance of $(\mathcal{P}1)$ is depicted, where NMSE is defined as $\mathbb{E}\left[\left\|\hat{\mathbf{h}}^{\mathrm{a}} - \mathbf{h}^{\mathrm{a}}\right\|_2^2 / \|\mathbf{h}^{\mathrm{a}}\|_2^2\right]$. The performance curves shown in this figure result from empirical simulations, where the non-zero elements of the channel vector $\mathbf{h}^{\mathrm{a}} \in \mathbb{C}^M$ are modeled as i.i.d. circularly-symmetric Gaussian variates with distribution $\mathcal{CN}(0, 1)$. To serve as a performance benchmark, a genie-aided least square (LS) estimator is considered, where perfect support knowledge at the UE side is exploited for channel recovery. Generally, for any given number N of linear measurements, the MSE decreases with increasing α. When N is lower, the performance gap between the curves of $\alpha = 0.8$ and $\alpha = 0.6$ becomes more significant. This highlights the importance of having a higher α, i.e., higher accuracy of partial support information. Note that the MSE performance due to $\alpha = 0.6$ is worse than that of using no prior information. It means that applying partial support information with low accuracy may not lead to performance improvement.

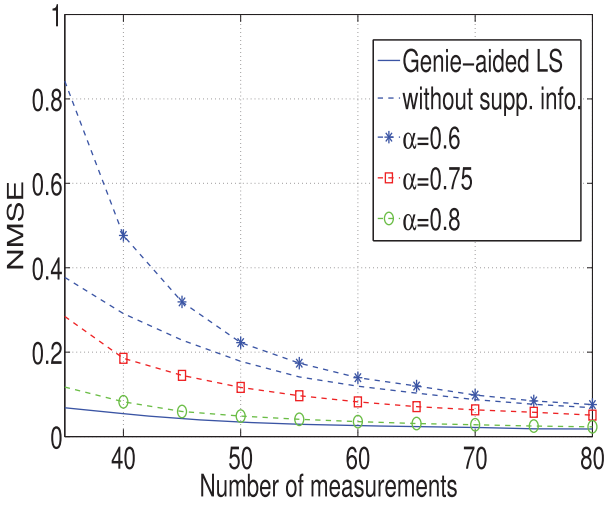


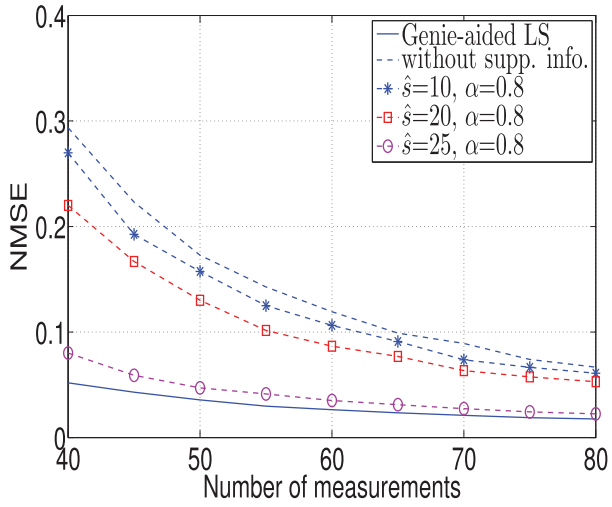
Fig. 3. Normalized mean square error versus number of measurements over different values of α given $M = 100$, $s = 20$, $\hat{s} = 25$, $\|\mathbf{e}\|_2 = \epsilon = 1$.

In Fig. 4, the MSE curves due to different values of $\hat{s}$ are compared when both the sparsity level s and the accuracy level α are fixed. It can be observed in Fig. 4a that having a higher $\hat{s}$ can lead to a lower MSE as $\alpha = 0.8$ is high enough. However, this improvement becomes less significant when the number of linear measurements gets larger. As shown in Fig. 4b, at a lower accuracy level $\alpha = 0.6$, the MSE performance of weighted ℓ_1 minimization is generally worse than that of normal ℓ_1 minimization without using any prior support knowledge. This result highlights that partial support information employed in weighted ℓ_1 minimization should be as accurate as possible.

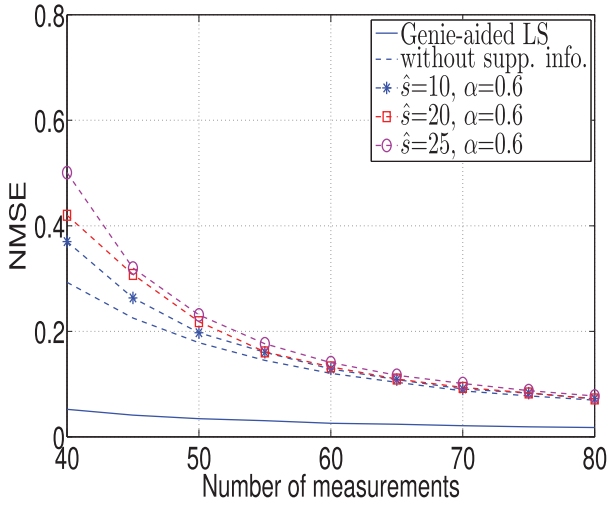
B. Phase Transition Characterization

Fig. 5 compares the analytical phase transition curves due to (23) where $N = \lfloor \omega(\Phi)^2 + 1/2 \rfloor$ with those empirical curves over different values of α. Basically, both analytical and empirical curves show that the required measurements to achieve a certain percentage of exact recovery increases gradually with an increasing sparsity level. We declare exact recovery if $||\hat{\mathbf{x}} - \mathbf{x}^*||_2 \leq 10^{-4}$ where $\mathbf{x}^*$ is modeled as a random vector with independent standard complex normal entries, and the value 10^{-4} is within the range suggested in [46, Appendix A.1]. As shown in this figure, the analytical curves of $\alpha = 0.2$ and $\alpha = 0.8$ can accurately depict the empirical phase transition curves of 60% exact recovery and 55% exact recovery, respectively. This observation demonstrates the capability of the proposed analytical result to characterize the phase transition of $(\mathcal{P}1)$ with partial support information. Fig. 6 shows how the training overhead changes with respect to a larger antenna array size. Although M is doubled from 100 to 200, the required training overhead for achieving 60% or 55% exact recovery does not increase proportionally.

Fig. 7 presents the probability of robust recovery as a function of the sparsity level and the number of linear measurements when there is some bounded error. The brightness corresponds to the recovery probability and robust recovery is declared if



(a) $\alpha = 0.8$



(b) $\alpha = 0.6$

Fig. 4. Normalized mean square error versus number of measurements given $M = 100$, $s = 20$, $\|\mathbf{e}\|_2 = \epsilon = 1$.

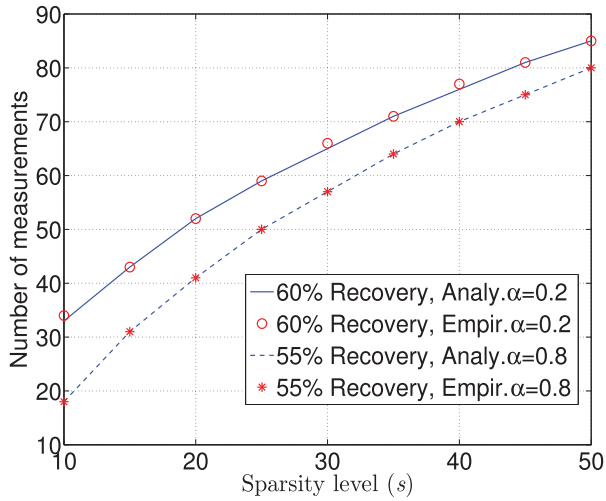


Fig. 5. Phase transition curves of ($\mathcal{P}1$) over different values of α given $M = 100$, $\hat{s} = 10$, $\mathbf{e} = \mathbf{0}$, and $\epsilon = 0$.

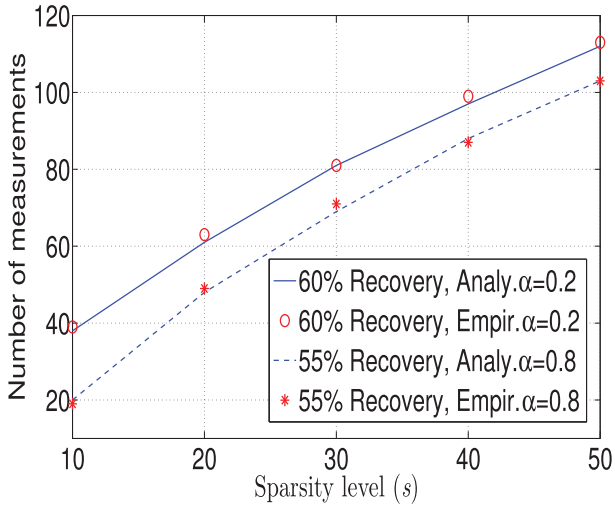


Fig. 6. Phase transition curves of ($\mathcal{P}1$) over different values of α given $M = 200$, $\hat{s} = 10$, $\mathbf{e} = \mathbf{0}$, and $\epsilon = 0$.

$||\hat{\mathbf{x}} - \mathbf{x}^*||_2 \leq 0.2$, where the value 0.2 is determined according to Theorem 5-2) once ϵ and ν are given. On top of this grayscale image, the empirical curves of 5%, 80%, and 90% robust recovery are plotted respectively. It can be seen that the analytical curve from (25) closely matches the empirical curve of 80% robust recovery. Meanwhile, the result in Corollary 6-2) can be further verified via this phase transition plot. Take the sparsity level $s = 10$ as an example. To reach 80% robust recovery, it requires 30 measurements as $\omega(\Phi)^2 \leq 57.4$ is considered in (25). Making use of $\beta_{\eta=0.8} = 1.8$ in (27) gives that the minimum required measurement size is 47, which is greater than the actually required size of 30. In other words, the bound provided by (27) is shown to be valid.

VI. Conclusions

This paper has addressed the issue of training overhead reduction in FDD massive MIMO by utilizing partial channel support information. It has been shown that a significant overhead reduction can be achieved via weighted ℓ_1 minimization which takes advantage of readily available partial support information. With this in mind, a novel analysis of the required training overhead has been developed by adopting a convex geometry approach. The numerical results validated the accuracy of our analytical characterization of the required training overhead.

The current investigation has examined the usage of partial support information within the framework of weighted ℓ_1 minimization. Future study could explore the possibility of having higher training overhead reduction within a more general reweighted ℓ_1 minimization framework [47].

While this study did not include the general case of multiuser channel recovery, it can be easily used to form the basis for such a scenario. In comparison to other single-user pilot training in [13], [14], the compressed sensing based pilot design is less user-specific. That means, our proposed approach is more suitable to be extended to address the case of multiuser channel recovery. A simple way to achieve this generalization is to

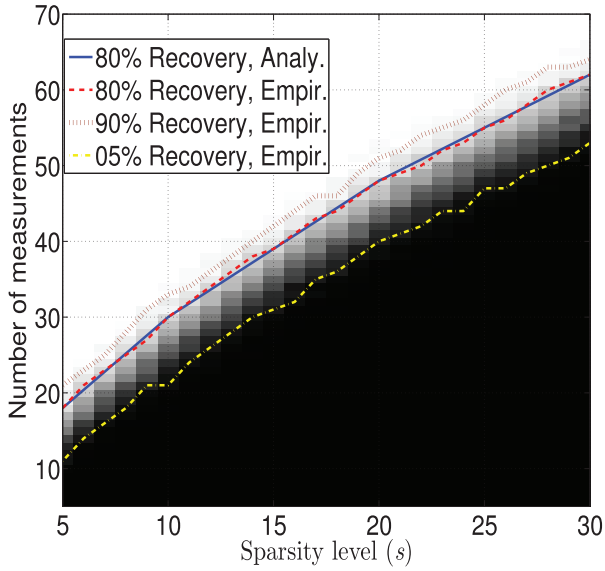


Fig. 7. Phase transition curves of ($\mathcal{P}1$) for different percentages of robust recovery given $M = 100$, $\hat{s} = 0$, $\|\mathbf{e}\|_2 = \epsilon = 10^{-3}$, $\nu = 10^{-2}$. The brightness corresponds to the recovery probability.

group together channels with the same sparsity level and SNR level, and learn this group of channels at the same time. This idea, similar to the user grouping proposed in [5], could be the subject of a future investigation.

Appendix A
Proof of Corollary 6

Here only the proof of (27) is provided. The second result of Theorem 5 holds with probability at least η if (24)$\geq \eta$, which can be easily shown to be equivalent to

$$\left(\lambda_{2N} - \omega(\Phi) - \sqrt{2N}\nu\right) \geq \beta_\eta. \tag{31}$$

The condition (31) is met when

$$\begin{aligned}\left(\lambda_{2N} - \sqrt{2N}\nu\right) &= \frac{2N}{\sqrt{2N+1}} - \sqrt{2N}\nu, \\ &\geq \frac{2N}{\sqrt{2N+1}} - \sqrt{2N+1}\nu, \\ &\geq \omega(\Phi) + \beta_\eta. \end{aligned} \tag{32}$$

We further recast the last inequality in (32) as

$$\begin{aligned}(1-\nu)^2(2N)^2 - \left[2\nu(1-\nu) + \left(\omega(\Phi) + \beta_\eta\right)^2\right](2N) \\ +\nu^2 - \left(\omega(\Phi) + \beta_\eta\right)^2 \geq 0,\end{aligned}$$

which is satisfied when

$$\begin{aligned}&2(2N)(1-\nu)^2 \\ &\geq \Bigg\{2\nu(1-\nu) + \left(\omega(\Phi) + \beta_\eta\right)^2 \\ &\quad + \Bigg\{\left[2\nu(1-\nu) + \left(\omega(\Phi) + \beta_\eta\right)^2\right]^2 \\ &\quad + 4(1-\nu)^2\left[\left(\omega(\Phi) + \beta_\eta\right)^2 - \nu^2\right]\Bigg\}^{1/2}\Bigg\}.\end{aligned} \tag{33}$$

The second term of the right-hand side of (33) is less than

$$\begin{aligned}&\Bigg\{\left[2\nu(1-\nu) + \left(\omega(\Phi) + \beta_\eta\right)^2\right]^2 \\ &\quad +4\left[\left(\omega(\Phi) + \beta_\eta\right)^2 + 2\nu(1-\nu)\right] + 4\Bigg\}^{1/2}.\end{aligned}$$

Therefore, we have the lower bound

$$\begin{aligned}2N &\geq \frac{\left(\omega(\Phi) + \beta_\eta\right)^2 + \frac{3}{2}}{(1-\nu)^2}, \\ &\geq \frac{2\left(\omega(\Phi) + \beta_\eta\right)^2 + 2\nu(1-\nu) + 2}{2(1-\nu)^2},\end{aligned} \tag{34}$$

where the second inequality holds due to $\nu(1-\nu) \leq 1/4$, and this bound satisfies the condition (33).

Appendix B
Proof of Lemma 7

Instead of presenting the weighted normal cone $\mathsf{N}_{2,1,\mathbf{w}}(\tilde{\mathbf{x}}^*)$, we derive the expression of the normal cone $\mathsf{N}_{2,1}(\tilde{\mathbf{x}}^*)$. The result can be easily extended to obtain the desired expression of $\mathsf{N}_{2,1,\mathbf{w}}$.

We first claim that the normal cone $\mathsf{N}_{2,1}(\tilde{\mathbf{x}}^*)$ at $\tilde{\mathbf{x}}^*$ with respect to $\mathcal{B}_{2,1}(\tilde{\mathbf{x}}^*)$ can be expressed as

$$\begin{aligned}\mathsf{N}_{2,1}(\tilde{\mathbf{x}}^*) = \Bigg\{\mathbf{v} \in \mathbb{R}^{2M} |\ v_i = t\frac{\tilde{x}_i^*}{|x_i^*|},\ v_{i+M} = t\frac{\tilde{x}_{i+M}^*}{|x_i^*|}, \\ \text{for } i \in T,\ \sqrt{v_i^2 + v_{i+M}^2} \leq t \text{ for } i \in T^c,\ \text{for } t \geq 0\Bigg\}.\end{aligned} \tag{35}$$

To verify the claim, we have the following derivation. By definition,

$$\begin{aligned}\mathsf{N}_{2,1}(\tilde{\mathbf{x}}^*) = \Big\{\mathbf{v} \in \mathbb{R}^{2M} |\ \langle \mathbf{v},\ \mathbf{z} - \tilde{\mathbf{x}}^*\rangle \leq 0 \\ \forall \mathbf{z} \text{ s.t. } \|\mathbf{z}\|_{2,1} \leq \|\tilde{\mathbf{x}}^*\|_{2,1}\Big\}.\end{aligned} \tag{36}$$

For any element $\mathbf{v}$ in (35), we have

$$\begin{aligned}&\langle \mathbf{v},\ \mathbf{z} - \tilde{\mathbf{x}}^*\rangle \\ &= \langle \mathbf{v},\ \mathbf{z}\rangle - \langle \mathbf{v},\ \tilde{\mathbf{x}}^*\rangle, \\ &= \sum_{i=1}^{M}(v_i z_i + v_{i+M} z_{i+M}) - \sum_{i\in T} t\left(\frac{\tilde{x}_i^*}{|x_i^*|}\tilde{x}_i^* + \frac{\tilde{x}_{i+M}^*}{|x_i^*|}\tilde{x}_{i+M}^*\right), \\ &\leq \sum_{i\in T} t\left(\frac{|\tilde{x}_i^*|}{|x_i^*|}|z_i| + \frac{|\tilde{x}_{i+M}^*|}{|x_i^*|}|z_{i+M}|\right) \\ &\quad + \sum_{i\in T^c} t\left(\frac{|v_i|}{t}|z_i| + \frac{|v_{i+M}|}{t}|z_{i+M}|\right) \\ &\quad - \sum_{i\in T} t\left[\frac{\left(\tilde{x}_i^*\right)^2 + \left(\tilde{x}_{i+M}^*\right)^2}{|x_i^*|}\right], \\ &\leq \sum_{i=1}^{M} t\sqrt{z_i^2 + z_{i+M}^2} - \sum_{i=1}^{M} t\sqrt{\left(\tilde{x}_i^*\right)^2 + \left(\tilde{x}_{i+M}^*\right)^2}, \\ &= t\left(\|\mathbf{z}\|_{2,1} - \|\tilde{\mathbf{x}}^*\|_{2,1}\right) \leq 0,\end{aligned} \tag{37}$$

where the Cauchy-Schwarz inequality is applied in the second inequality, and the last inequality is due to the mixed-norm inequality in (36). The inequality (37) implies that $\mathbf{v}$ is also in $\mathsf{N}_{2,1}(\mathbf{x}^*)$.

On the other hand, if $\mathbf{v}$ is in (36), then

$$\begin{aligned}
&\sup_{\mathbf{z}} \left\{ \langle \mathbf{v},\ \mathbf{z}-\tilde{\mathbf{x}}^* \rangle \mid \|\mathbf{z}\|_{2,1} \le \|\tilde{\mathbf{x}}^*\|_{2,1} \right\} \\
&= \sup_{\mathbf{z}} \left\{ \langle \mathbf{v},\ \mathbf{z} \rangle \mid \|\mathbf{z}\|_{2,1} \le \|\tilde{\mathbf{x}}^*\|_{2,1} \right\} - \langle \mathbf{v},\ \tilde{\mathbf{x}}^* \rangle, \\
&= \|\tilde{\mathbf{x}}^*\|_{2,1} \sup_{\bar{\mathbf{z}}} \left\{ \langle \mathbf{v},\ \bar{\mathbf{z}} \rangle \mid \|\bar{\mathbf{z}}\|_{2,1} \le 1,\ \bar{\mathbf{z}} = \frac{\mathbf{z}}{\|\tilde{\mathbf{x}}^*\|_{2,1}} \right\} \\
&\quad - \langle \mathbf{v},\ \tilde{\mathbf{x}}^* \rangle, \\
&= \|\tilde{\mathbf{x}}^*\|_{2,1}\, v_{\max} - \langle \mathbf{v},\ \tilde{\mathbf{x}}^* \rangle, \qquad (38)
\end{aligned}$$

where $v_{\max} = \max\left\{ \sqrt{v_i^2 + v_{i+M}^2},\ 1 \le i \le M \right\}$ and the definition of the norm dual to the mixed $\ell_{2,1}$ norm is applied in the last equality. As (38) is less than or equal to zero, we have

$$\begin{aligned}
&\|\tilde{\mathbf{x}}^*\|_{2,1} \\
&\le \frac{1}{v_{\max}} \langle \mathbf{v},\ \tilde{\mathbf{x}}^* \rangle, \\
&= \sum_{i \in T} \left(\frac{v_i}{v_{\max}} \tilde{x}_i^* + \frac{v_{i+M}}{v_{\max}} \tilde{x}_{i+M}^* \right), \\
&\le \sum_{i \in T} \left(\frac{|v_i|}{\sqrt{v_i^2 + v_{i+M}^2}} |\tilde{x}_i^*| + \frac{|v_{i+M}|}{\sqrt{v_i^2 + v_{i+M}^2}} |\tilde{x}_{i+M}^*| \right), \\
&\le \|\tilde{\mathbf{x}}^*\|_{2,1}, \qquad (39)
\end{aligned}$$

where $v_{\max} > 0$ is assumed. All inequalities in (39) become exact equalities, leading to $v_i = v_{\max} \tilde{x}_i^* / |x_i^*|$ and $v_{i+M} = v_{\max} \tilde{x}_{i+M}^* / |x_i^*|$ for $i \in T$. So $\mathbf{v}$ is also in (35) for $v_{\max} \ge 0$.

Appendix C
Proof of Proposition 8

The square of Gaussian width $\omega(\Phi)$ is given by

$$\begin{aligned}
\omega(\Phi)^2 &= \omega\left(\mathcal{T}_{2,1,\mathbf{w}}(\tilde{\mathbf{x}}^*) \cap \mathbb{S}^{2M-1} \right)^2, \\
&\le \mathbb{E}_{\mathbf{g}} \left[\operatorname{dist}\left(\mathbf{g},\ \mathsf{N}_{2,1,\mathbf{w}}(\tilde{\mathbf{x}}^*) \right) \right]^2, \\
&\le \mathbb{E}_{\mathbf{g}} \left[\operatorname{dist}\left(\mathbf{g},\ \mathsf{N}_{2,1,\mathbf{w}}(\tilde{\mathbf{x}}^*) \right)^2 \right], \\
&= \mathbb{E}_{\mathbf{g}} \left[\inf_{\mathbf{v} \in \mathsf{N}_{2,1,\mathbf{w}}(\tilde{\mathbf{x}}^*)} \|\mathbf{g}-\mathbf{v}\|_2^2 \right], \qquad (40)
\end{aligned}$$

where dist[4] represents the Euclidean distance between a point and a set, the first inequality follows from [43, Prop. 3.6], and Jensen's inequality is applied to the second inequality.

[4]The Euclidean distance between the point $\mathbf{x}$ and the set A is given by $\operatorname{dist}(\mathbf{x}, A) = \inf_{\mathbf{y} \in A} \|\mathbf{x}-\mathbf{y}\|_2$.

Making use of Lemma 7 yields

$$\begin{aligned}
&\inf_{\mathbf{v} \in \mathsf{N}_{2,1,\mathbf{w}}(\tilde{\mathbf{x}}^*)} \|\mathbf{g}-\mathbf{v}\|_2^2 \\
&= \inf_{\substack{t \ge 0 \\ |v_i| \le t \text{ for } i \in T^c \setminus \widehat{T}}} \sum_{i \in \widehat{T}} \left[g_i^2 + g_{i+M}^2 \right] \\
&\quad + \sum_{i \in T \setminus \widehat{T}} \left[\left(g_i - t \frac{\tilde{x}_{i+M}^*}{|x_i^*|} \right)^2 + \left(g_{i+M} - t \frac{\tilde{x}_{i+M}^*}{|x_i^*|} \right)^2 \right] \\
&\quad + \sum_{i \in T^c \setminus \widehat{T}} \left[(g_i - v_i)^2 + (g_{i+M} - v_{i+M})^2 \right], \\
&= \inf_{t \ge 0} \sum_{i \in \widehat{T}} \left[g_i^2 + g_{i+M}^2 \right] \\
&\quad + \sum_{i \in T \setminus \widehat{T}} \left[g_i^2 + g_{i+M}^2 - 2t \left(\frac{g_i \tilde{x}_i^*}{|x_i^*|} + \frac{g_{i+M} \tilde{x}_{i+M}^*}{|x_i^*|} \right) + t^2 \right] \\
&\quad + \sum_{i \in T^c \setminus \widehat{T}} \operatorname{Pos}^2\left(\sqrt{g_i^2 + g_{i+M}^2} - t \right), \qquad (41)
\end{aligned}$$

where

$$\operatorname{Pos}\left(\sqrt{g_i^2 + g_{i+M}^2} - t \right) = \max\left\{ 0,\ \left(\sqrt{g_i^2 + g_{i+M}^2} - t \right) \right\}.$$

Hence,

$$\begin{aligned}
&\mathbb{E}_{\mathbf{g}} \left[\inf_{\mathbf{v} \in \mathsf{N}_{2,1,\mathbf{w}}(\tilde{\mathbf{x}}^*)} \|\mathbf{g}-\mathbf{v}\|_2^2 \right] \\
&= \inf_{t \ge 0} \left\{ \sum_{i \in \widehat{T}} \mathbb{E}_{\mathbf{g}} \left[g_i^2 + g_{i+M}^2 \right] \right. \\
&\quad + \sum_{i \in T \setminus \widehat{T}} \mathbb{E}_{\mathbf{g}} \left[g_i^2 + g_{i+M}^2 - 2t \left(\frac{g_i \tilde{x}_i^*}{|x_i^*|} + \frac{g_{i+M} \tilde{x}_{i+M}^*}{|x_i^*|} \right) + t^2 \right] \\
&\quad \left. + \sum_{i \in T^c \setminus \widehat{T}} \mathbb{E}_{\mathbf{g}} \left[\operatorname{Pos}^2\left(\sqrt{g_i^2 + g_{i+M}^2} - t \right) \right] \right\}, \\
&= \inf_{t \ge 0} \left\{ 2 \cdot \operatorname{card}\left(\widehat{T} \right) + \left(2 + t^2 \right) \cdot \operatorname{card}\left(T \setminus \widehat{T} \right) \right. \\
&\quad \left. + \operatorname{card}\left(T^c \setminus \widehat{T} \right) \cdot \int_t^\infty (r-t)^2\, r \exp\left(-\frac{r^2}{2} \right) \mathrm{d}r \right\}, \\
&= \inf_{t \ge 0} \left\{ 2\hat{s} + (s - \alpha\hat{s}) \left(2 + t^2 \right) \right. \\
&\quad \left. + [M - \hat{s} - (s - \alpha\hat{s})] \int_t^\infty (r-t)^2\, r \exp\left(-\frac{r^2}{2} \right) \mathrm{d}r \right\}, \qquad (42)
\end{aligned}$$

where the following equality is applied

$$\begin{aligned}&\mathbb{E}_{\mathbf{g}}\left[\mathrm{Pos}^2\left(\sqrt{g_i^2+g_{i+M}^2}-t\right)\right]\\&=\mathbb{E}_r\left[\mathrm{Pos}^2\left(r-t\right)\right],\\&=\int_t^{\infty}(r-t)^2\, r\exp\left(-\frac{r^2}{2}\right)\mathrm{d}r,\end{aligned}\tag{43}$$

in which $\sqrt{g_i^2+g_{i+M}^2}$ is replaced by a Rayleigh distributed variable r.

The way to show that t_{opt} is the unique minimizer of the right-hand side of (29) is similar to that in [46, Prop. 4.5].

REFERENCES

[1] J.-C. Shen, J. Zhang, E. Alsusa, and K. B. Letaief, "Compressed CSI acquisition in FDD massive MIMO with partial support information," in *Proc. IEEE Int. Conf. Commun.*, London, U.K., Jun. 2015, pp. 1459–1464.

[2] 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.

[3] J.-C. Shen, J. Zhang, K.-C. Chen, and K. B. Letaief, "High-dimensional CSI acquisition in massive MIMO: Sparsity-inspired approaches," *IEEE Syst. J.*, 2016.

[4] J. Jose, A. Ashikhmin, T. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[5] 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.

[6] K. Pedersen, P. Mogensen, and B. Fleury, "Power azimuth spectrum in outdoor environments," *Electron. Lett.*, vol. 33, no. 18, pp. 1583–1584, Aug. 1997.

[7] D.-S. Shiu, G. Foschini, M. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502–513, Mar. 2000.

[8] F. Adachi, M. Feeney, J. Parsons, and A. Williamson, "Crosscorrelation between the envelopes of 900 MHz signals received at a mobile radio base station site," *IEE Proc. F Commun. Radar Signal Process.*, vol. SP-133, no. 6, pp. 506–512, Oct. 1986.

[9] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, Feb. 2013.

[10] S. Wu, C.-X. Wang, H. Haas, E.-H. Aggoune, 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.

[11] S. Payami and F. Tufvesson, "Channel measurements and analysis for very large array systems at 2.6 GHz," in *Proc. Eur. Conf. Antennas Propag. (EUCAP)*, Prague, Czech Republic, Mar. 2012, pp. 433–437.

[12] X. Gao, F. Tufvesson, O. Edfors, and F. Rusek, "Measured propagation characteristics for very-large MIMO at 2.6 GHz," in *Proc. Asilomar Conf. Signals Syst. Comp. (ASILOMAR)*, Pacific Grove, CA, USA, Nov. 2012, pp. 295–299.

[13] J. H. Kotecha and A. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 546–557, Feb. 2004.

[14] E. Bjornson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated Rician MIMO channels with Rician disturbance," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1807–1820, Mar. 2010.

[15] E. Bjornson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7112–7139, Nov. 2014.

[16] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[17] L. Zheng and D. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

[18] M. Kobayashi, G. Caire, and N. Jindal, "How much training and feedback are needed in MIMO broadcast channels?" in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Toronto, ON, Canada, Jul. 2008, pp. 2663–2667.

[19] J. Choi, D. 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.

[20] X. Rao and V. 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.

[21] X. Rao and V. 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.

[22] C. Qi, Y. Huang, S. Jin, and L. Wu, "Sparse channel estimation based on compressed sensing for massive MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, U.K., Jun. 2015, pp. 4558–4563.

[23] W. Shen, B. Wang, J. Feng, C. Gao, and J. Ma, "Differential CSIT acquisition based on compressive sensing for FDD massive MIMO systems," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, Glasgow, U.K., May 2015, pp. 1–4.

[24] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

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

[26] R. Ertel, P. Cardieri, K. Sowerby, T. Rappaport, and J. Reed, "Overview of spatial channel models for antenna array communication systems," *IEEE Pers. Commun.*, vol. 5, no. 1, pp. 10–22, Feb. 1998.

[27] O. Besson and P. Stoica, "Decoupled estimation of DOA and angular spread for a spatially distributed source," *IEEE Trans. Signal Process.*, vol. 48, no. 7, pp. 1872–1882, Jul. 2000.

[28] A. Rao and D. Jones, "Efficient detection with arrays in the presence of angular spreading," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 301–312, Feb. 2003.

[29] K. Hugl, K. Kalliola, and J. Laurila, "Spatial reciprocity of uplink and downlink radio channels in FDD systems," COST 273 Tech. Document, vol. 66, Espoo, Finland, May 2002.

[30] Y.-C. Liang and F. Chin, "Downlink channel covariance matrix (DCCM) estimation and its applications in wireless DS-CDMA systems," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 2, pp. 222–232, Feb. 2001.

[31] J. Foutz, A. Spanias, and M. K. Banavar, *Narrowband Direction of Arrival Estimation for Antenna Arrays*. San Rafael, CA, USA: Morgan & Claypool, 2008.

[32] G. Xu and H. Liu, "An effective transmission beamforming scheme for frequency-division-duplex digital wireless communication systems," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Detroit, MI, USA, May 1995, pp. 1729–1732.

[33] J. Yin and T. Chen, "Direction-of-arrival estimation using a sparse representation of array covariance vectors," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4489–4493, Sep. 2011.

[34] Z.-M. Liu and Y.-Y. Zhou, "A unified framework and sparse Bayesian perspective for direction-of-arrival estimation in the presence of array imperfections," *IEEE Trans. Signal Process.*, vol. 61, no. 15, pp. 3786–3798, Aug. 2013.

[35] M. Souden, S. Affes, and J. Benesty, "A two-stage approach to estimate the angles of arrival and the angular spreads of locally scattered sources," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1968–1983, May 2008.

[36] 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.

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

[38] C. Herzet, C. Soussen, J. Idier, and R. Gribonval, "Exact recovery conditions for sparse representations with partial support information," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7509–7524, Nov. 2013.

[39] P.-H. Kuo, H. 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)*, Shanghai, China, Apr. 2012, pp. 492–497.

[40] M. A. Davenport, P. T. Boufounos, and R. G. Baraniuk, "Compressive domain interference cancellation," *SPARS'09 Signal Processing with Adaptive Sparse Structured Representations*, Saint Malo, France, Apr. 2009.

[41] Z. Yang, C. Zhang, and L. Xie, "On phase transition of compressed sensing in the complex domain," *IEEE Signal Process. Lett.*, vol. 19, no. 1, pp. 47–50, Jan. 2012.
[42] D. Donoho and J. Tanner, "Precise undersampling theorems," *Proc. IEEE*, vol. 98, no. 6, pp. 913–924, Jun. 2010.
[43] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Found. Comput. Math.*, vol. 12, no. 6, pp. 805–849, 2012.
[44] Y. Gordon, "On Milman's inequality and random subspaces which escape through a mesh in $\mathbb{R}^n$," in *Geometric Aspects of Functional Analysis*. New York, NY, USA: Springer, 1988, vol. 1317, pp. 84–106.
[45] A. Maleki, L. Anitori, Z. Yang, and R. Baraniuk, "Asymptotic analysis of complex LASSO via complex approximate message passing (CAMP)," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4290–4308, Jul. 2013.
[46] D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp, "Living on the edge: A geometric theory of phase transitions in convex optimization," *Inf. Inference*, vol. 3, no. 3, pp. 224–294, 2014.
[47] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, nos. 5–6, pp. 877–905, 2008.

Juei-Chin Shen (S'10–M'14) received the Ph.D. degree in communication engineering from the University of Manchester, Manchester, U.K., in 2013. He was with the Hong Kong University of Science and Technology (HKUST), Clear Water Bay, Hong Kong, from 2013 to 2015, as a Research Associate. Since 2015, he has been with Mediatek Inc., as a Senior Engineer, working on the research and standardization of future wireless technologies. His research interests include compressed sensing, massive MIMO systems, and millimeter-wave communications. He was the corecipient of IEEE PIMRC 2014 Best Paper Award.

Jun Zhang (S'06–M'10–SM'15) received the B.Eng. degree in electronic engineering from the University of Science and Technology of China, Hefei, China, in 2004, the M.Phil. degree in information engineering from the Chinese University of Hong Kong, Hong Kong, in 2006, and the Ph.D. degree in electrical and computer engineering from the University of Texas at Austin, Austin, TX, USA, in 2009. He is currently a Research Assistant Professor with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology (HKUST), Clear Water Bay, Hong Kong. He has coauthored the book *Fundamentals of LTE* (Prentice-Hall, 2010). His research interests include wireless communications and networking, green communications, and signal processing. He is an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and served as a MAC track Co-Chair for the IEEE WCNC 2011. He was the recipient of the 2014 Best Paper Award for the *EURASIP Journal on Advances in Signal Processing*, and the PIMRC 2014 Best Paper Award.

Emad Alsusa (M'06–SM'07) received the Ph.D. degree in electrical and electronic engineering from Bath University, Bath, U.K., in 2000.

He then joined the School of Engineering and Electronics, Edinburgh University, Edinburgh, U.K., as a MobileVCE Postdoctoral Research Fellow, working on link enhancement techniques for future high-data rate wireless communication systems. In 2003, he joined the University of Manchester, Manchester, U.K., as an Academic Member of the School of Electrical and Electronic Engineering, where he lectures on communication engineering subjects. His research interests include signal processing techniques and analysis of wireless communication networks, with particular focus on cognitive radio, interference mitigation, multiuser MIMO, GreenComm, and energy and spectrum optimization techniques. He has served as a Technical Program Committee member on numerous IEEE flagship conferences and chaired the Manchester EEE postgraduate conference in 2010.

Khaled B. Letaief (S'85–M'86–SM'97–F'03) received the B.S. degree with *distinction*, M.S. and Ph.D. degrees in electrical engineering from Purdue University at West Lafayette, Indiana, USA. From 1990 to 1993, he was a faculty member at the University of Melbourne, Australia. He has been with the Hong Kong University of Science and Technology since 1993, where he is known as one of HKUST's most distinguished professors for his boundless energy, collegial nature, dedication, and excellence in research, education, and service. While at HKUST, he has held numerous administrative positions, including Chair Professor and Dean of HKUST School of Engineering, Head of the Electronic and Computer Engineering department, Director of the Center for Wireless IC Design, Director of Huawei Innovation Laboratory, and Director of the Hong Kong Telecom Institute of Information Technology.

From September 2015, he joined HBKU as Provost to help establish a research-intensive university in Qatar in partnership with strategic partners that include Northwestern, CMU, Cornell, and Texas A&M. He is an internationally recognized leader in wireless communications and networks and served as consultants for different organizations including Huawei, ASTRI, ZTE, Nortel, PricewaterhouseCoopers, and Motorola. He is the founding Editor-in-Chief of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and has served on the editorial board of other prestigious journals. He has also been involved in organizing a number of flagship international conferences and events.

In addition to his active research and professional activities, Professor Letaief has been a dedicated teacher committed to excellence in teaching and scholarship. He received the *Mangoon Teaching Award* from Purdue University in 1990; the Teaching Excellence Appreciation Award by the School of Engineering at HKUST (4 times); and the Michael G. Gale Medal for Distinguished Teaching (*Highest university-wide teaching award* and only one recipient/year is honored for his/her contributions).

He is also the recipient of many other distinguished awards and honors including the 2007 IEEE Communications Society *Joseph LoCicero* Publications Exemplary Award, 2009 IEEE Marconi Prize Award in Wireless Communications, 2010 Purdue University Outstanding Electrical and Computer Engineer Award, 2011 IEEE Communications Society Harold Sobol Award, 2011 IEEE Wireless Communications Technical Committee Recognition Award, and 12 IEEE Best Paper Awards.

Dr. Letaief is recognized as a long time volunteer with dedicated service to professional societies and in particular IEEE where he has served in many leadership positions. These include *Treasurer* of the IEEE Communications Society, Vice-President for Conferences of the IEEE Communications Society, Vice-President for Technical Activities of the IEEE Communications Society, Chair of the IEEE Committee on Wireless Communications, member of the IEEE Fellow Evaluation Committee, and elected member of the IEEE Product Services and Publications Board.

Dr. Letaief is a Fellow of IEEE and a Fellow of HKIE. He is currently serving as the IEEE Communications Society Director of Journal, and member of IEEE TAB Periodicals Committee. He is also recognized by Thomson Reuters as an *ISI Highly Cited Researcher*.