CHANNEL ESTIMATION FOR MASSIVE MIMO USING SPARSE BAYESIAN LEARNING METHOD

[1]M.Keerthi, [2]T.Ravi Babu, [3]Manas Ranjan Biswal

[1]Student Scholar M.Tech, [2,3]Assistant Professor
Electronics and Communication Engineering,
Sanketika Vidya Parishad Engineering College, Visakhapatnam, India

***Abstract*: Pilot contamination creates a fundamental limit to the potential benefits on the performance of massive multiple-input multiple-output (MIMO) systems. Due to failure in accurate channel estimation the author proposed spares Bayesian learning (SBL) method on Gaussian framework.. To address this problem, we propose to estimate channel coefficient by its own hyper parameter and also to its adjacent cells. The required estimation is, nonetheless, an underdetermined system. In this paper the simulation results show that the channel coefficients can be estimated more efficiently in contrast to the conventional channel estimators in terms of channel estimation with pilot contamination. A pilot design criterion is proposed to design the optimal pilot to improve the estimation accuracy of the proposed algorithm using the Lagrange multiplier optimization method. Results show that we can reduce the MSE of the SBL estimator by employing the optimal pilot sequence. As a result, if the signals are observed in the beam domain (using Fourier transform), the channel is approximately sparse, i.e., the channel matrix contains only a small fraction of large components, and other components are close to zero.**

***Index Terms*: Spares Bayesian learning, Channel estimation, Gaussian Process, Massive MIMO, OFDM Modulation, Pilot contamination.**

I. INTRODUCTION

Very large multiple-input multiple-output (MIMO) or "massive MIMO" systems [1] are widely considered as a future cellular network architecture, which are anticipated to be energy-efficient, spectrum-efficient, secure, and robust for a survey. Such systems employ a few hundred or more base station (BS) antennas to simultaneously serve many tens of user equipment's (UEs) in the same radio channel. As such, the array gain is expected to grow unboundedly with the number of antennas at the BSs so that both multiuser interference and thermal noise for any given number of users and any given powers of the interfering users can be eliminated.

The reports on the great benefits of massive MIMO systems, however, were based on the assumption that the BSs have an acceptable quality of channel knowledge, which in practice has to be estimated via finite-length pilot sequences. However, in cellular networks, pilot interference from neighboring cells limits the ability to obtain sufficiently accurate channel estimates, giving rise to the problem of "pilot contamination". It was noted that pilot contamination [1] incurs an ultimate limit on the interference rejection performance on massive MIMO, even if the number of antennas grows without bound.

In this paper, our focus is on the channel estimation problems with pilot contamination in the uplink, although there are other related issues in the downlink that also greatly limit the performance of massive MIMO systems. For the issues in the downlink, we refer the readers to Several approaches have emerged to deal with pilot contamination in the uplink recently [10-15]. By exploiting the covariance information of user channels and applying a covariance-aware pilot assignment strategy among the cells [10], revealed that pilot contamination could disappear. Alternatively, using an eigenvalue decomposition of the sample covariance matrix of the received signals, claimed that pilot contamination can be effectively mitigated by projecting the received signal [11-13] onto an interference free subspace without the need of coordination amongst the cells. Nevertheless, rely heavily on the estimation of the channel or signal covariance matrices. Though the covariance matrices change slowly over time, the estimation problem under massive MIMO systems is far from trivial [5]. The reason is that a covariance matrix is typically estimated through the sample covariance matrix, and that the sample size should be increased proportionally to the dimension of the covariance matrices. In massive MIMO systems, the dimension of the covariance matrices may be comparable to the number of available samples within a coherence time. The sample covariance estimation method is thus no longer sufficient and more sophisticated techniques must be used. Different from the approaches based on covariance matrices, in this paper, we address the pilot contamination problem directly from a channel estimation perspective. We realize that pilot contamination results from performing channel estimation ignoring pilot interference from the neighboring cells so that the estimated channel contains channels of the interference. To overcome this, we therefore propose to estimate not only the channel parameters of the desired links in the target cell but also those of the interference links from adjacent cells. Although this strategy seems natural, the challenge remains that the required estimation problem forms an underdetermined linear system which generally has infinitely many solutions.

To get an accurate solution, we rely on a key observation–The channels with most of the multipath energy tend to be concentrated in relatively small regions within the channel angular spread due to limited local scatterers at the BSs. An approximate sparsity of a channel can be obtained by transforming the received signal into a beam domain. Exploiting the channel sparsity, we can obtain much more accurate channel estimates by leveraging on more recent techniques in compressive sensing. MIMO channel estimation based on CS techniques has been investigated. Most of the earlier works, exploited sparse channel estimation methods mainly to improve the performance of single-user MIMO systems. Under multiuser massive MIMO systems, CS techniques were used in order to reduce the feedback overhead of the channel state information (CSI) at the transmitter side. the authors also advocated to estimate the channel parameters of the desired links in the target cell and those of the interference links from adjacent cells. Nonetheless, they used a CS technique to estimate the MIMO channel based on low-rank approximation,

which is completely different from that of our interest. Other popular solvers in the CS literature, e.g., the ℓ1 optimization (L1) solver and the orthogonal matching pursuit (OMP) solver, also appear to be not so useful in the concerned channel estimation problem. For the L1 solver, the regularization parameter has to be chosen carefully to control the channel estimation errors while determining the best regularization parameter is difficult in practice. Meanwhile, t he OMP solver greedily selects the best channel vectors for channel representation, and the best support number for channel representation is also difficult to obtain in practice. Whet her channel estimation in massive MIMO systems, suffered from pilot contamination, could be effectively addressed via CS techniques is not understood.

A. The formulation of massive MIMO channel estimation with pilot contamination:

Our contributions include the formulation of massive MIMO channel estimation with pilot contamination as a CS problem. Based on an observation of the received signals in the beam domain, we model the channel component in the beam domain as a Gaussian mixture, i.e., a weighted summation of Gaussian distributions with different variances. This model enables us to reconstruct the channel components based on the probabilistic Bayesian inference with the best mean-squared error (MSE) performance. For the optimal Bayesian inference, the computational complexity is not tractable and the statistical properties of the channel component are required. Hence, we employ the approximate message passing (AMP) algorithm to obtain the Bayesian inference and an expectation - maximization (EM) algorithm to learn the statistical properties. Unlike our Bayesian estimator does not require the availability of the channel covariance matrices and the background noise level. All the required channel knowledge will be learned as part of the estimation procedure. By a proper design on pilot sequences, the proposed estimator leads to a much reduced complexity without compromising performance. Numerical results will show that the developed approach pro-vides a huge gain in reducing the channel estimation errors. In addition, the achievable rates based on the developed channel estimator are comparable to those with perfect CSI.

Notations:

Throughout this paper, the set of complex numbers is denoted by C. For any matrix $A \in C^{M\times N}$, A_{ij} denotes the (i , j) th element, while A^T , and A^H return the transpose and the conjugate transpose of A, respectively. An identity matrix is denoted by I or I_N if it is necessary to specify its dimension N. In addition, a random vector x having the proper complex Gaussian distribution of mean μ and covariance Ω is indicated by $x \sim N_C(x;\mu,\Omega)$, where

$$N_C(x;\mu,\Omega)=\frac{1}{\det(\pi\Omega)}e^{-(x-\mu)^H\Omega^{-1}(x-\mu)} \quad (1)$$

We simply denote $N_C(0;\mu,\Omega)$ by $N_C(\mu,\Omega)$ for conciseness. Finally, $E\{\cdot\}$ returns the expectation of an input random entity.

B. Introduction to Approximate Message Passing:

Approximate message passing (AMP) and its variants are a powerful class of algorithms for linear inverse problems and their generalizations. AMP methods were originally developed for compressed sensing problems of estimating sparse vectors from underdetermined linear measurements. They have now been extended to a wide range of estimation and learning problems including regularized least squares, generalized linear models, matrix completion, dictionary learning and estimation in networks of systems with linear and nonlinear blocks. The key appealing features of the methods are their computational scalability and generality. In addition, in certain large random instances, the performance of the methods can be precisely characterized with testable conditions for Bayes optimality, even in non-convex instances.

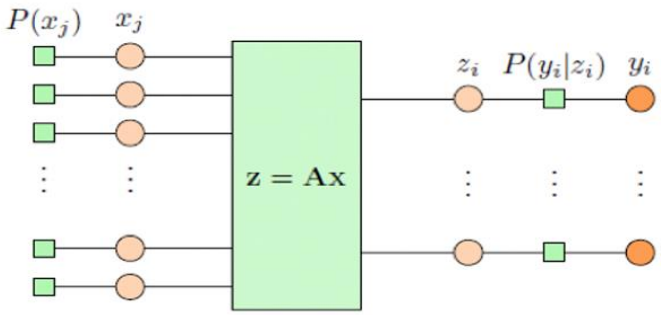

Fig.1 Block diagram for AMP

C. Introduction to Least Mean Squares (LMS) Algorithms:

The least mean squares (LMS) algorithms adjust the filter coefficients to minimize the cost function. Compared to recursive least squares (RLS) algorithms, the LMS algorithms do not involve any matrix operations. Therefore, the LMS algorithms require fewer computational resources and memory than the RLS algorithms. The implementation of the LMS algorithms also is less complicated than the RLS algorithms. However, the eigenvalue spread of the input correlation matrix, or the correlation matrix of the input signal, might affect the convergence speed of the resulting adaptive filter.

D. Introduction to Recursive Least Square (RLS) Algorithm:

The Recursive least squares (RLS) adaptive filter is an algorithm which recursively finds the filter coefficients that minimize a weighted linear least squares cost function relating to the input signals. The RLS algorithms are known for their excellent performance when working in time varying environments but at the cost of an increased computational complexity and some stability problems. In this algorithm the filter tap weight vector is updated using Equations.

w(n) = wT (n-1) + k(n) en-1(n) ..... (a)
k(n) = u(n) / (λ+X T (n) u(n)) .....(b)
u(n) = wλ -1 (n-1) X(n) . .....(c)

II. PROBLEM DEFINETION AND EXISTING METHODS

A. Recursive Least Square (RLS) Algorithm:

The recursive least square error (RLS) filter is a sample adaptive, time-update, version of the Wiener filter. For stationary signals, the RLS filter converges to the same optimal filter coefficients as the Wiener filter. For nonstationary signals, the RLS filter tracks the time variations of the process. The RLS filter has a relatively fast rate of convergence to the optimal filter coefficients. This is useful in applications such as speech enhancement, channel equalization, echo cancellation and radar where the filter should be able to track relatively fast changes in the signal process.

Unlike the LMS filter, who updates the coefficients by single parameter μ, the recursive least square algorithm perform this update by a vector k(m) called the Kalman gain vector and the corresponding equation is given by

$$w(n) = w(n+1) + k(n)e(n) \quad (2)$$

Where

$$k(n) = \frac{M_{rr}^{-1}(n-1)r(n)}{1+r^T(n)M_{rr}^{-1}(n-1)r(n)} \quad (3)$$

And

$$M_{rr}(n) = \sum_{n=0}^{N-1} r(n)r^T(n) \quad (4)$$

Is the autocorrelation Matrix of the noisy signal r(n).

B. Normalized LMS:

Fast Block LMS:

Some adaptive filter applications, such as adaptive echo cancellation and adaptive noise cancellation, require adaptive filters with a large filter length. If you apply the standard LMS algorithm to the adaptive filter, this algorithm might take a long time to complete the filtering and coefficients updating process. This length of time might cause problems in these applications because the adaptive filter must work in real time to filter the input signals. In this situation, you can use the fast block LMS algorithm.

The fast block LMS algorithm uses the fast Fourier transform (FFT) to transform the input signal x(n) to the frequency domain. This algorithm also updates the filter coefficients in the frequency domain. Updating the filter coefficients in the frequency domain can save computational resources. The fast block LMS algorithm differs from the standard LMS algorithm in the following ways:

- The fast block LMS algorithm updates the coefficients of an adaptive filter block by block. The block size is exactly the same as the filter length. However, the standard LMS algorithm updates the filter coefficients sample by sample.
- The fast block LMS algorithm requires fewer multiplications than the standard LMS algorithm. If both the filter length and block size are N, the standard LMS algorithm requires N(2N+1) multiplications, whereas the fast block LMS algorithm requires only (10Nlog2N+26N) multiplications. If N = 1024, the fast block LMS algorithm can execute 16 times faster than the standard LMS algorithm.

The fast block LMS algorithm calculates the output signal and the error signal before updating the filter coefficients. The following diagram illustrates the steps that this algorithm completes to calculate these signals.

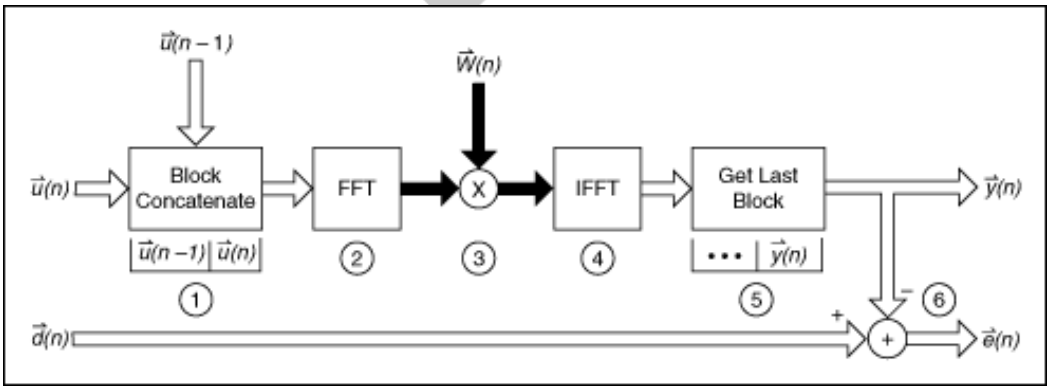

Fig.2 Block diagram of fast block LMS

In the previous figure, the fast block LMS algorithm completes the following steps to calculate the output and error signals.

1. Concatenates the current input signal block to the previous blocks.
2. Performs an FFT to transform the input signal blocks from the time domain to the frequency domain.
3. Multiplies the input signal blocks by the filter coefficients vector.

4. Performs an inverse FFT (IFFT) on the multiplication result.
5. Retrieves the last block from the result as the output signal vector.
6. Calculates the error signal vector by comparing the input signal vector with.

After calculating the output and error signals, the fast block LMS algorithm updates the filter coefficients. The following diagram shows the steps that this algorithm completes to update the filter coefficients.

III. SYSTEM MODEL

A. MASSIVE MIMO:

Consider a wireless communication system with B cells, in which each cell contains a BS and K UEs. Each BS has N antennas, whereas each UE is equipped with a single antenna. In the considered uplink training phase, all UEs in the B cells simultaneously transmit pilot sequences of length T symbols. For ease of exposition, we let the first cell be our target cell . The pilot sequences used in the bth cell can be represented by a T × K matrix, S_b, and the corresponding channel vector between the UEs in the bth cell and the target BS is denoted by $H_b = [h_{b1} \cdots h_{bK}]^T \in C^{K\times N}$, where $h_{bk} \in C^{N\times 1}$ is the channel from UE_k in cell b to the target BS. The received signals during uplink training at the target BS is written as

$$Y = \sum_{b=1}^{B} S_b H_b + Z \square SH + Z. \quad (5)$$

where $Z \in C^{T\times N}$ denotes the temporally and spatially white Gaussian noise with zero mean and element-wise variance Δ. Also in (5), we have defined $S \square [S_1 .....S_B] \in \square^{T\times BK}$ and $H \square [H_1^H .....H_B^H]^H \in \square^{BK\times N}$ for conciseness.

B. PILOT CONTAMINATION:

In massive MIMO, the statistical knowledge of the channel matrix would be practically unknown because the size of the channel matrix would mean that an unacceptably large number of samples would be required. In this case, the standard way of estimating H is to employ the least square (LS) approach. If orthogonal pilot sequences are adopted in the bth cell, i.e., $S^H_b S_b = I_K$, and the same pilot sequences are reused in all B cells, i.e., $S_1 = \cdots = S_B$, the outputs of the LS estimator at the targeted BS can be written as

$$\hat{H}_1 = (S_1^H S_1)^{-1} S_1 Y = H_1 + \sum_{b\neq 1}^{B} H_b + (S_1^H S_1)^{-1} S_1 Z. \quad (6)$$

From the perspective of the LS estimator, the assumption of using the same set of pilot sequences makes no fundamental difference in terms of estimation performance compared with using different pilots in different cells [1]. Clearly, in (6), the interfering channels will leak directly to the desired channel estimate, which gives rise to "pilot contamination".

The fundamental effect of pilot contamination can also be understood from other perspective through linear estimation theory. First, we note that if the BK × N channel matrix **H** can be estimated from the T × N measurement matrix Y with sufficient accuracy, then the pilot contamination [1,4,10] effect can be mitigated or eliminated. A straightforward requirement for an accurate channel estimation is T ≥ BK; otherwise, unknown variables will outnumber measurements and in this case accurate channel estimation is clearly impossible. Un-fortunately, the requirement for accurate channel estimation usually cannot be satisfied in the massive MIMO system because most scenarios of our interests have T ≈ K and B > 1.[2] The estimation of H from the noisy underdetermined measurement has infinitely many solutions. For this reason, many speculate that the pilot contamination problem will exist regardless of which channel estimation method is used [1,4]. Clearly, to get a correct solution, one must impose extra constraints in choosing the solution. In the next section, we will reveal that such additional constraint do exist, thanks to the propagation properties of the massive MIMO system.

IV. BEAM DOMAIN CHANNEL ESTIMATION

A. Sparsity Characteristics:

In a typical cellular configuration, the channel from a UE to a BS is a correlated random vector with a covariance matrix that depends on the scattering geometry [5]. Following, the channel vectors $h^T_{bk} \in C^{1\times N}$ can be modeled by

$$h_{bk}^T = v_{bk}^T R_{bk}^{\frac{1}{2}} \quad (7)$$

where R_{bk} denotes a positive semi-definite channel covariance matrix and $v_{bk} \sim N_C(v_{bk}; 0, I_N)$. Let F denote the N × N discrete Fourier transform (DFT) matrix. Taking the DFT of the channel vectors h^T_{bk} leads to

$$h_{bk}^T = h_{bk}^T F = v_{bk}^T R_{bk}^{\frac{1}{2}} F. \quad (8)$$

We refer to h_{bk}^T as the beam domain channel representation of h_{bk}^T . The nth element of h_{bk}^T corresponds to the channel response observed at the nth beam. The most crucial property of h_{bk}^T is that the elements of h_{bk}^T is approximately sparse, i.e., the channel vector contains only a small fraction of large components, and the other components are close to zero.

The sparsity property stated above can be easily realized by the argument as follows: Consider that the BSs are equipped with a uniform linear array (ULA) of half wavelength spacing. With an infinite number of antennas at the BSs, the DFT matrix R_{bk} [5]. We illustrate the argument by an example, the covariance matrix is generated by

$$R_{bk} = \int_A p(\theta)\, a(\theta)\, a^H(\theta) d\theta, \quad (9)$$

Where $a(\theta) \square [1, e^{-j\pi\sin\theta}, ..., e^{-j\pi(N-1)\sin\theta}]^T$ is the ULA steering vector [34], $p(\theta)$ denotes the power azimuth spread (PAS), and $A = (-\pi/2, \pi/2)$ represents the angle of arrival (AOA) region see at the BS. Following the argument of [5,20], there is a one-to-one mapping between θ and $n \in \{0,1,....., N-1\}$ such that $\sin\theta - (n/N) \to 0$ as $N \to \infty$. Therefore, $a(\theta)$ can be served as a DFT basis and $p(\theta)$ is the corresponding eigenvalue. Let $R_{bk} = F\Lambda_{bk}F^H$, where Λ_{bk} is the diagonal matrix whose diagonal elements are the corresponding eigenvalues $\{p(n/N)\}_{n=0,1,....N-1}$. In typical outdoor propagations, the PAS can be modeled by a Laplacian distribution is given by

$$p(\theta) = \frac{1}{\sqrt{2}\sigma_{AS}} e^{-\frac{\sqrt{2}|\theta-\bar{\theta}|}{\sigma_{AS}}}, \quad (10)$$

where θ corresponds to the mean AOA of the UE channel and σ_{AS} denotes the azimuth spread (AS). An example of p(θ) with θ = 0 and $\sigma_{AS} = 3$ is illustrated in Figure 1(a). As we see from the figure, most of the eigenvalues are (nearly) zero. Then we infer from (4.4) and $R_{bk} = F\Lambda_{bk}F^H$ that

$$\underline{h}_{bk}^T = v_{bk}^T F \Lambda_{bk}^{\frac{1}{2}} \quad (11)$$

is sparse.

Though in a practical setting, the numbers of antennas at the BSs are finite but large, **F** still can be an approximate eigen vector matrix of R_{bk} .Following N=256, the corresponding channel magnitude in the beam domain $|\underline{h}_{bk}^T|$ is depicted. As is expected, the channel vector is not perfectly sparse but approximately sparse. Specifically, over 99% of the channel total power is located only within about 16% of the beam indices. The channel magnitude of the approximately sparse components highly depends on the number of antennas. The larger the number of antennas the better a(θ) matching to the DFT basis. Moreover, its sparsity property is related to the PAS of the channel model. Despite the fact that laplacian distributions is the most popular model for the PAS. There are other c lasses of distribution which serve as better models under certain circumstances. From many experimental measurements of MIMO channels, it is believed that as the number of antenna increases , the channel responses in the beam domain tend to be sparse due to the limited number of local scatterers at the BS.

Taking the DFT of Y we can therefore obtain the received signal in the beam domain given by

$$\underline{Y} = S\underline{H} + \underline{Z}, \quad (12)$$

Where $\underline{Y} \square YF$, $\underline{H} \square HF$, and $\underline{Z} \square ZF$. Because the DFT matrix is a unitary matrix, the statistical property of $\underline{Z}$ is the same as that of **Z.** Thus, the difference between Y and $\underline{Y}$ is only at the channel matrix. In contrast to H in Y, $\underline{H}$ in $\underline{Y}$ is approximately sparse. To gain an idea on $\underline{H}$, we generate $B \times K = 4 \times 80$ UEs whose PAS follow the Laplacian distribution (12) with $\sigma_{AS} = 3^\circ$ and $\bar{\theta}$ being the uniformly distributed random variable within $[-90^\circ, 90^\circ)$. The corresponding pseudeo color plot of the strength of $\underline{H} \in \square^{320\times250}$ is depicted in. As expected, the channel matrix observed in the beam domain $\underline{H}$ is approximately sparse.

For ease of expression, we use $\underline{y}_n \in \square^T$, $s_n \in \square^T$, $\underline{h}_n \in \square^{BK}$, and $\underline{Z}$ respectively. With the definitions, the received signals in the nth beam domain can be read

$$\underline{y}_n = S\underline{h}_n + \underline{z}_n \quad (13)$$

We aim to estimate $\underline{h}_n$ based on $\underline{y}_n$ given the full knowledge of the pilot matrix **S.** Note that to get S, we should acquire the pilot sequence of the desired links and those of the adjacent cells. In the undetermined system of interest, the pilot sequence are no longer orthogonal and thus are randomly generated. To proceed with the estimation process for each n, we set the element-wise variance of $\underline{z}_n$ as Δ_n even though we may have $\Delta_n = \Delta.\forall_n$.

Before proceeding, we find it useful to see a picture on $\underline{h}_n$. Note that the row index of $\underline{H}$ corresponds to the beam index observed at the BS. Thus, the nth column of $\underline{H}$ represents the channel responses of the whole UEs observed at the nth beam. A realization of the real parts of $\underline{h}_n$ at n = 140 is depicted. As can be seen from the figure, the elements of $\underline{h}_n = [\underline{h}_{k,n}]$ contain only a small fraction of large components and the other components are close to zero. In addition, they seem to be statistically independent. Another important observation is that the interference links from adjacent cells also appear to have strong channels, which are the main source of serve pilot contamination.

It is evident that the conventional LS estimator will not be able to address such undetermined systems so that the strong interfering channels will leak to the desired channel estimate. However, taking advantage of the approximate sparsity of $\underline{h}_n$, we can use CS to obtain a sufficiently accurate estimate of the channel responses from the undetermined system. CS can be thought of as a technique that can automatically focus on the estimate of the stronger channel responses while treating very weak channel

responses as noise. As such, the underdetermined system reduces to the determined system and the pilot contamination effect can hence be mitigated.

B. Sparsity Characteristics:

Among various CS approaches, probabilistic Bayesian inference has recently attracted much attention for its outstanding recovery performance. In order to apply probabilistic Bayesian inference to (13), one requires to know the distribution of $\underline{h}_n$. To this end, the following two observations are useful. First, it can infer from (12) that each element of $\underline{h}_n$ consists of a Gaussian random variable, although one should particularly notice that $\underline{h}_n$ and $\underline{h}_{bk}^T$ in (12) are observed from different perspective. Second, we observe that the elements of $\underline{h}_n$ have significantly different variances, i.e., some of them are very small but some are large.

Inspired by the two observations, we model the elements of $\underline{h} = [\underline{h}_{n,k}]$ by a Gaussian-mixture (GM) distribution:

$$p(\underline{h}_{k,n}; \rho_n, \sigma_n^2) = \sum_{l=1}^{L} \rho_{n,l} N_{\square}(\underline{h}_{k,n}; 0, \sigma_{n,l}^2), \tag{14}$$

where $N_{\square}(\underline{h}_{k,n}; 0, \sigma_{n,l}^2)$ denotes a Gaussian probability density function (pdf) with zero mean and variance $\sigma_{n,l}^2$, and $\rho_{n,l}$, is the mixing probability of the l^{th} GM component. The parameter $\sigma_{n,l}^2$ can be set to a very small value so that $\rho_{n,l}$ denotes the density of the components close to zero. The remaining GM components $\{\rho_{n,l}, \sigma_{n,l}^2\}_{l=2}^{L}$ can be used to model the small fraction of large components. The value of L reflects the number of different variances in . We will discuss the setting of L in Section IV-B. Note that the true distributions of $\underline{h}_n$ could not be the GM distribution. However, our numerical results will demonstrate that the choice of the GM distribution is perfectly fine. Finally, we assume that the BK-dimensional $\underline{h}_n$ contains independent and identically distributed (i.i.d.) components, so we have

$$P(\underline{h}_n; \eta_n) = \prod_{k=1}^{BK} P(\underline{h}_{k,n}; \eta_n) \quad with \quad \eta_n \square \left(\rho_n, \sigma_n^2\right). \tag{15}$$

For conciseness, we often omit η_n from $P(\underline{h}_n; \eta_n)$, and the index n from $\rho_{n,l}, \sigma_{n,l}^2, \rho_n, \sigma_n, \eta_n$.

As an example the empirical pdf based on the channel responses in Both the estimated GM and Gaussian distributions are provided. The parameters are obtained from a learning algorithm to be described later while the variance of the Gaussian distribution is the sample variance. Clearly, the GM distribution provides a significantly better fit than the Gaussian distribution.

The Bayes-optimal way to estimate $\underline{h}_n$ that minimizes the MSE is given by

$$\hat{\underline{h}}_{k,n} = \int \underline{h}_{k,n} Q(\underline{h}_{k,n}) d\underline{h}_{k,n} \tag{16}$$

Where

$$Q(\underline{h}_{k,n}) = \int P(\underline{h}_n \mid \underline{y}_n) \prod_{i \neq k} d\underline{h}_{i,n} \tag{17}$$

denotes the marginal pdf of the k^{th} variable under the posterior measure . From Bayes theorem, the posterior distribution can be written as $P(\underline{h}_n \mid \underline{y}_n) = P(\underline{y}_n \mid \underline{h}_n) P(\underline{h}_n) / P(\underline{y}_n)$, Where the conditional distribution of $\underline{y}_n$ based on (10) reads $P(\underline{y}_n \mid \underline{h}_n) = \frac{1}{(\pi\Delta_n)^T} e^{-\frac{1}{\Delta_n}\left\|\underline{y}_n - S\underline{h}_n\right\|^2}$.

There are two issues when implementing the optimal Bayes estimation (17). First, the optimal Bayes estimation (17) is not computationally tractable. Second, the prior parameters η_n are unknown. To obtain an estimate of the marginal pdfs $\{Q(\underline{h}_{k,n})\}$, we adopt the AMP algorithm in which is an iterative message passing algorithm. Meanwhile, we use the EM algorithm to learn the prior parameters η_n. We describe the two algorithms and their connections in the remaining part of this subsection.

We begin with the AMP algorithm. Suppose that the elements of $\underline{h}_n$ follows the GM distribution with prior parameters η_n. Then the updating rules for estimating $Q(\underline{h}_{k,n})$ are given as (for the derivation with these notations, see :

$$V_{m,n}^{t+1} = \sum_j \left|S_{mj}\right|^2 \upsilon_{j,n}^t, \tag{18a}$$

$$\omega_{m,n}^{t+1} = \sum_j S_{mj} a_{j,n}^t - \frac{V_{m,n}^{t+1}}{\Delta_n + V_{m,n}^t} (\underline{y}_{m,n} - \omega_{m,n}^t), \tag{18b}$$

$$(\Sigma_{k,n}^2)^{t+1} = [\sum_i \frac{\left|S_{ik}\right|^2}{\Delta_n + V_{i,n}^{t+1}}]^{-1}, \tag{18c}$$

$$R_{k,n}^{t+1} = a_{k,n}^{t} + (\sum_{k,n}^{2})^{t+1} \sum_{i} \frac{S_{ik}^{*}(\underline{y}_{i,n} - \omega_{i,n}^{t+1})}{\Delta_n + V_{i,n}^{t+1}}, \tag{18d}$$

$$a_{k,n}^{t+1} = f_a((\sum_{k,n}^{2})^{t+1}, R_{k,n}^{t+1}; \eta_n), \tag{18e}$$

$$\upsilon_{k,n}^{t+1} = f_c((\sum_{k,n}^{2})^{t+1}, R_{k,n}^{t+1}; \eta_n), \tag{18f}$$

where t = 0, 1, . . . represents the iteration index. Here, f_a and f_c are some analytical functions depending on $P(\underline{h}_{k,n})$ and will be given in (17) later in this subsection

AMP employs central limit theorem to approximate sum of many random variables as a Gaussian. To get an intuition on the algorithm, we provide an interpretation on each step of (16) while we refer the interested readers to for detailed derivations. First of all, we view the target estimate $\hat{\underline{h}}_{k,n}$ as a Gaussian with mean $a_{k,n}^{t}$ and variance $v_{k,n}^{t}$ at the t^{th} iteration. Therefore, if we temporarily ignore the second term of (18b), $\omega_{m,n}^{t+1}$ can be understood as a current mean estimate of the m^{th} element of $S\underline{h}_n$, which has variance $V_{m,n}^{t+1}$ given by (18a). Next, considering the conventional estimator by the matched filter, we get $S^H \underline{y}_n \approx \hat{\underline{h}}_n^t + S^H(\underline{y}_n - S\hat{\underline{h}}_n^t)$, the approximation follows from the fact that $S^H S \approx I$. Then (18d) can be read as the k^{th} element of

$$\hat{\underline{h}}_n^t + noise = [R_{k,n}^{t+1}], \tag{19}$$

where noise above is Gaussian distributed with zero mean and variance of $(\sum_{k,n}^{2})^{t+1}$ for its k^{th} element. Clearly, $R_{k,n}^{t+1}$ in (18d) and $(\sum_{k,n}^{2})^{t+1}$ in (18c) are the mean and variance of the current estimate of $\underline{h}_{k,n}$ without taking into account the prior information of $\underline{h}_{k,n}$. Finally, in (18e)–(18f), we estimate the mean and the variance of $\underline{h}_n^{t+1}$ from (19) by taking into account the prior information of $\underline{h}_{k,n}$. Specifically, from (19), the posterior probability of $\underline{h}_{k,n}$ after the observation of $R_{k,n}^{t+1}$ is given by

$$\begin{aligned} P(\underline{h}_{k,n} \mid R_{k.n}) &= \frac{P(R_{k,n} \mid \underline{h}_{k,n}) P(\underline{h}_{k,n})}{\int P(R_{k,n} \mid \underline{h}_{k,n}) P(\underline{h}_{k,n})\, d\underline{h}_{k,n}} \\ &= \frac{N_{\square}(R_{k,n}; \underline{h}_{k,n}, \sum_{k,n}^{2}) P(\underline{h}_{k,n})}{\int N_{\square}(R_{k,n}; \underline{h}_{k,n}, \sum_{k,n}^{2}) P(\underline{h}_{k,n})\, d\underline{h}_{k,n}}, \end{aligned} \tag{20}$$

Where $R_{k.n}$ and $\sum_{k,n}^{2}$ change from one iteration to another while we have suppressed t for brevity. Plugging the GM prior $P(\underline{h}_{k,n})$ (15) into (23), we get

$$\begin{aligned} &P(\underline{h}_{k,n} \mid R_{k,n}) \\ &= \sum_{l=1}^{L} \hat{\rho}_{k,n,l} N_{\square}\left(\underline{h}_{k,n}; \frac{\sigma_{n,l}^2 R_{k,n}}{\sigma_{n,l}^2 + \sum_{k,n}^{2}}, \frac{\sigma_{n,l}^2 \sum_{k,n}^{2}}{\sigma_{n,l}^2 + \sum_{k,n}^{2}}\right), \end{aligned} \tag{21}$$

Where $\hat{\rho}_{k,n,l} \square \frac{\rho_{n,l} N_{\square}\left(R_{k,n}, \sigma_{n,l}^2 + \sum_{k,n}^{2}\right)}{\sum_{l=1}^{L} \rho_{n,l} N_{\square}\left(R_{k,n}, \sigma_{n,l}^2 + \sum_{k,n}^{2}\right)}$. As a consequence the posterior mean variance of $\underline{h}_{k,n}$ are respectively given by $f_a\left(\sum_{k,n}^{2}, R_{k,n}; \eta_n\right)$ and $f_c\left(\sum_{k,n}^{2}, R_{k,n}; \eta_n\right)$, where

$$f_a\left(\sum^2, R; \eta\right) \square \frac{\sum_{l=1}^{L} \rho_l N_{\square}\left(R, \sum^2 + \sigma_l^2\right) \frac{R\sigma_l^2}{\sum^2 + \sigma_l^2}}{\sum_{l=1}^{L} \rho_l N_{\square}\left(R, \sum^2 + \sigma_l^2\right)}, \tag{22a}$$

$$f_b\left(\sum^2, R; \eta\right) \square \frac{\sum_{l=1}^{L} \rho_l N_{\square}\left(R, \sum^2 + \sigma_l^2\right) \frac{\sigma_l^2 \sum^2 (\sum^2 + \sigma_l^2) + |R|^2 \sigma_l^4}{\sum^2 + \sigma_l^2}}{\sum_{l=1}^{L} \rho_l N_{\square}\left(R, \sum^2 + \sigma_l^2\right)}, \tag{22b}$$

$$f_c\left(\sum^2, R; \eta\right) \square f_b\left(\sum^2, R; \eta\right) - \left| f_a\left(\sum^2, R; \eta\right) \right|^2. \tag{22c}$$

The procedures (20) in conjunction with f_a and f_c in (21) lead to the AMP algorithm. After the convergence of the iterative equations (20) is established, we get the estimated marginal pdf $Q(\underline{h}_{k,n})$, which is approximated by a Gaussian distribution with mean $a_{k,n}^t$ and variance $\upsilon_{k,n}^t$. Hence, the posterior mean estimate of $\underline{h}_{k,n}$ is obtained as $a_{k,n}^t$.

Before proceeding, let us discuss the computational complexity for performing (18) in the massive MIMO application. First, notice that the iterative equations are general formulas for the sum and product of BK (or T) variables. In fact, a significant computational saving can be obtained if the pilot are designed as sequences of $\{\pm\sqrt{\xi}\}$, where ξ is chosen to fulfill the power constraint, e.g., if $\|s_k\|^2 = 1$ then ξ = 1/T . In this case, one can effectively replace every $|s_{ij}|^2$ by ξ in (18a) and (18c). Thus, $V_{m,n}^{t+1}$ and $(\sum_{k,n}^2)^{t+1}$ are independent on the indices m and k. Therefore, we can define $V_n^{t+1} = V_{m,n}^{t+1}$, ∀m and $(\sum_n^2)^{t+1} = (\sum_{k,n}^2)^{t+1}$, ∀k so that (18a) and (18c) can take a much simpler form as shown in Step 3 and Step 5 of Algorithm 1, respectively. Now, the two steps only involve addition operations. No multiplication is required. In addition, (18b) and (18d) also take the simpler form as those in Steps 4 and 6 of Algorithm 1, respectively. In the steps, the product between Sik and any number can work with a sign bit operation. The whole operation procedures are summarized in Algorithm 1. Note that except for the second term in Step 4, where one multiplication is required for each m index (m = 1, . . . ,T), all the other steps (i.e., Steps 3–6) can be implemented only through additions. In addition, Steps 7 and 8 involve sum and product of L terms for each k index (k = 1, . . . , B K). In summary, the complexity of Algorithm 1 grows as O(T) + O(BK) per iteration. The complexity of Algorithm 1 is indeed higher than that of the LS estimator.6 However, LS would not work in the underdetermined system to have acceptable performance.

V. SIMULATION RESULTS

A. Bayesian Estimator without Pilot Contamination:

In the previous subsections, computer simulations were presented under an artificial scenario, where the channel responses were generated directly from a GM distribution. Now, we provide simulation results to demonstrate the capability of the estimator in Algorithm 2 under the typical channel model defined at the beginning of Section IV. In this realistic case, we do not possess the true statistical knowledge of the channel responses. The true distributions of the channel responses are unknown and could not be the GM distribution. In the simulation, we considered the system with B = 1 (no pilot contamination) for various numbers of antennas at the BS

$$N \in \{128, 256, 512\}. \tag{23}$$

The Bayesian estimator uses different GM order L ∈ {2, 3, 5} and operates under the case of random pilot sequences. Figure3 displays the MSE versus the SNR. Note that the MSE defined in (27) has been normalized for different numbers of

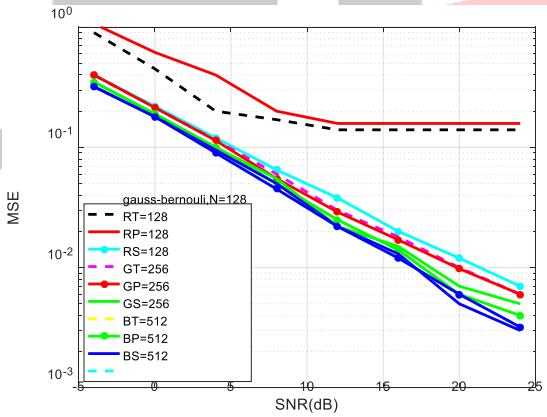

Fig.3. The MSE versus the system SNR for the Bayesian estimator with B = 1 (no pilot contamination) under different numbers of antennas at the BS.

UEs and antennas. The normalized MSEs are compared under different numbers of antennas at the BS.

It can be observed from Figure4 that when the GM order L increases, the MSE of the Baysian estimator decreases. Increasing L from 3 to 5 only gives minor improvement. In fact, through detailed numerical studies, we find that the choice of L = 3 is well enough. In the same figure, we also show the MSE of the Baysian estimator when the prior of $\underline{h}_{k,n}$ is assumed to be Gauss-Bernoulli (GB),i.e.,

$$P(\underline{h}_{k,n}) = (1-\tau_{n,1})\delta(\underline{h}_{k,n}) + \tau_{n,1} N_{\square}(\underline{h}_{k,n}; 0, \sigma_{n,1}^2). \tag{24}$$

GB distribution has been widely adopted in prior works where the channels are modeled to be perfectly sparse instead of approximately sparse. The EM algorithm was also used to learn the prior parameters of (24). Clearly, the Baysian estimator

with the GB prior cannot work well. Finally, we observe that the MSE of the Baysian estimator decreases when we increase the number of antennas N. This is because when the number of antennas increases, the steering vector a(θ) becomes the DFT basis and therefore the channel responses in the beam domain tend to have the smaller $\sigma_{n,1}^2$. According to the discussion presented in Section IV-A, the smaller the $\sigma_{n,1}^2$, the lower the MSE at high SNR.

B. Bayesian Estimator under Pilot Contamination:

Next, we fix N = 256 and repeat the previous simulations with B set to 4. In this case, channel estimation is performed under an underdetermined system. The regularized LS (R-LS) estimator

$$\hat{H}_1 = \left(S_1^H S_1 + \alpha I \right)^{-1} S_1 Y \tag{25}$$

is used to provide a performance reference, where α is the variance of the interference in the second term of (2). We assume that α is available in the R-LS estimator. Orthogonal pilot sequences are adopted in the R-LS estimator while random pilots are employed in the Baysian estimator.

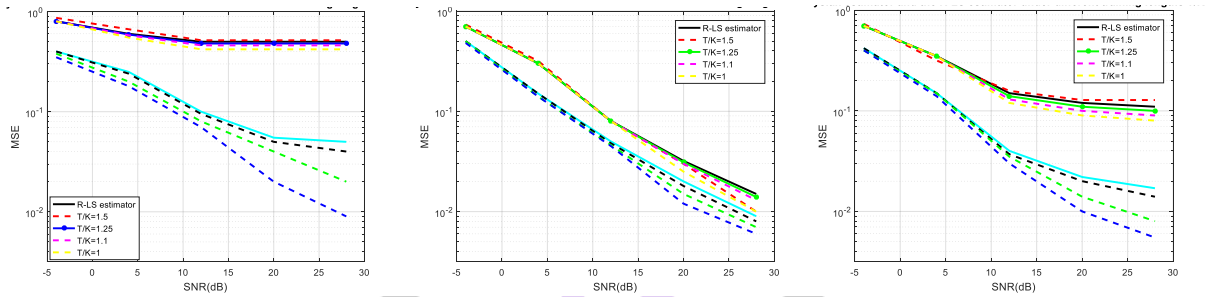

Fig.4. The MSE versus system SNR for the Baysian estimator and the R-LS estimator under different training lengths T/K = 1, 1.1, 1.25, 1.5. a)B′ = B = 4 for D = 1, b) B′ =B = 4 for D = 2, and c) B′ = 2 for D = 4.

PROPOSED SIMULATION RESULTS:

In the simulation, we considered the system with B = 1 (no pilot contamination) for various numbers of antennas at the BS

$$N \in \{128, 256, 512\}.$$

The Bayesian estimator uses different GM order L ∈ {2, 3, 5} and operates under the case of random pilot sequences. Figure 5 displays the MSE versus the SNR. Note that the MSE has been normalized for different numbers of UEs and antennas. The normalized MSEs are compared under different numbers of antennas at the BS.

It can be observed from Figure 6 that when the GM order L increases, the MSE of the Baysian estimator decreases. Increasing L from 3 to 5 only gives minor improvement. In fact, through detailed numerical studies, we find that the choice of L = 3 is well enough.

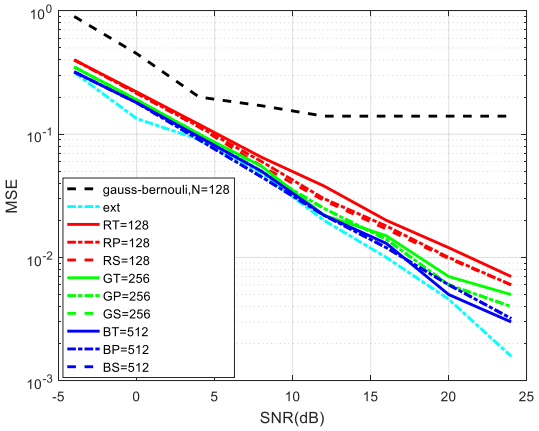

Fig.5.The MSE versus the system SNR for the Bayesian estimator with B = 1 (no pilot contamination) under different numbers of antennas at the Base station.

Clearly, the Baysian estimator with the GB prior cannot work well. Finally, we observe that the MSE of the Baysian estimator decreases when we increase the number of antennas N. This is because when the number of antennas increases.

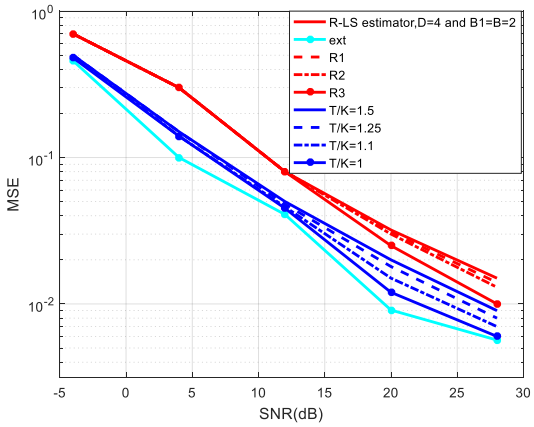

Fig.6.The MSE versus the system SNR for the Bayesian estimator with B = 1 (with pilot contamination) under different numbers of antennas at the Base station.

Next, we fix N = 256 and repeat the previous simulations with B set to 4. In this case, channel estimation is performed under an underdetermined system. Orthogonal pilot sequences are adopted in the R-LS estimator while random pilots are employed in the Baysian estimator.

Figure 6 shows the channel estimation error under different interfering localization distances, i.e., the distance between an interfering UE and the target BS is given by d = D+U , where D = 1, 2, or 4 and U ∈ (0, 1). Recall that we have d < 1 for the desired users in the target cell. The settings of D = 1, 2, 4 would be close to a cell plan with the reuse factors of 1, 3 and 7, respectively. In the simulations, we used L = 3.

Finally, following the same settings as those of Figure 4, Figure 5 depicts the corresponding average user rate achievable by the MRC receivers. As a reference, we show the average user rate of the MRC receiver based on perfect channel knowledge. We see that the Baysian estimator shows the average user rates comparable to the perfect channel knowledge and provides significant gain over the R-LS estimator. We know that the pilot contamination resulting in poor user rate is because the estimated channel contains channels of strong interference from neighboring cells. Thus, the results of this figure indicate that the Baysian estimator presents a substantial decontamination in terms of the strong interference.

VI. CONCLUSION

To address pilot contamination in massive MIMO systems, we proposed to estimate not only the channel parameters of the desired links in a target cell, but also that of the interference links from the adjacent cells. The channel estimation problem constitutes an underdetermined system. By transforming the received signals into the beam domain, we showed that the channel estimation problem can be solved using sparse Bayesian learning techniques. For the Bayesian approach, a good knowledge about the statistical properties of the channels is required. We modeled the channel component in the beam domain as a GM distribution and used EM to learn the prior parameters. Simulation results revealed that GM is much fine r than the conventional GB distribution in CS. In addition, to make the optimal Bayes estimation tractable, we employed the AMP algorithm, and a significant computational saving was obtained by designing the pilots appropriately. The proposed channel estimation approach does not require the availability of the channel covariance matrices, the background noise level, nor the need for coordination amongst the cells. Results illustrated that the developed channel estimator presents a substantial improvement over the conventional estimators in the presence of pilot contamination.

REFERENCES

[1] 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.

[2] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, “Scaling up MIMO: opportunities and challenges with very large arrays,” IEEE Signal Proces. Mag., vol. 30, no. 1, pp. 40–46, Jan. 2013.

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

[4] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, “The multicell multiuser MIMO uplink with very large antenna arrays and a finite-dimensional channel,” IEEE Trans. Wireless Commun., vol. 61, no. 6, pp. 2350–2361, Jun. 2013.

[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] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, “Massive MU-MIMO downlink TDD systems with linear precoding and downlink pilots,” in Allerton Conf. Communication, Control, and Computing, Urbana-Champaign, Illinois, Oct. 2013, pp. 293–298.

[7] 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 Sig. Process., vol. 8, no. 5, pp. 802–814, Oct. 2014.

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

[9] Y. Shi, J. Zhang, and K. B. Letaief, "CSI overhead reduction with stochastic beamforming for cloud radio access networks," in Proc. IEEE Int. Conf. Communications (ICC), Sydney, Australia, Jun. 2014, pp. 5154–5159.

[10] 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.

[11] H. Q. Ngo and E. G. Larsson, "EVD-based channel estimations for multicell multiuser MIMO with very large antenna arrays," in Proc. IEEE Int. Conf. Acoustics, Speed and Signal Processing (ICASSP), Kyoto, Japan, Mar. 2012, pp. 3249–3252.

[12] R. M¨uller, M. Vehkaper¨a, and L. Cottatellucci, "Blind pilot decontamination," in Proc. ITG Workshop on Smart Antennas, Stuttgart, Mar. 2013, pp. 1–6.

[13] R. M¨ueller, L. Cottatellucci, and M. Vehkaper¨a, "Blind pilot decontamination," IEEE J. Sel. Sig. Process., vol. 8, no. 5, pp. 773–786, Oct. 2014.

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

[15] "Precoding for multicell MIMO systems with compressive rankq channel approximation," in Proc. IEEE Personal, Indoor and Mobile Radio Communications, (PIMRC'03), London, United Kingdom, Sep. 2013, pp. 1227–1232.

[16] O. Ledoit and M. Wolf, "A well conditioned estimator for largedimensional covariance matrices," Journal of Multivariate Analysis, vol. 88, pp. 365–411, 2004.

[17] "Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions," 2013. [Online]. Available: http://dx.doi.org/10.2139/ssrn.2198287

[18] L. M. Correia, Wireless Flexible Personalized Communications. Wiley, 1st ed., 2001.

[19] S. Saunders and A. Arag´on-Zavala, Antennas and propagation for wireless communication systems. John Wiley & Sons, 2007.

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

[21] S. Payami and F. Tufvesson, "Channel measurements and analysis for very large array systems at 2.6 Ghz," in Proc. 6th European Conf.