Bayesian Compressed Sensing-based Channel Estimation for Massive MIMO Systems

Hayder Al-Salihi and Mohammad Reza Nakhai
Centre for Telecommunications Research
Department of Informatics, King's College London
London, UK
Email: ana.al-salihi@kcl.ac.uk

Abstract—The efficient and highly accurate channel state information (CSI) at the base station is essential to achieve the potential benefits of massive multiple input multiple output (MIMO) orthogonal frequency division multiplexing (OFDM) systems, due to limitations of the pilot contamination problem. It has recently been shown that compressed sensing (CS) techniques can address the pilot contamination problem, however, the CS-based channel estimation requires prior knowledge of channel sparsity. To solve this problem, in this paper, an efficient channel estimation approach based on Bayesian compressed sensing (BCS) that based on prior knowledge of statistical information about the channel sparsity is therefore proposed for the uplink of multi-user massive MIMO systems. Simulation results show that the proposed method can reconstruct the original channel coefficient effectively when compared to conventional based channel estimation.

Index Terms—Massive MIMO, Bayesian Compressed Sensing (BCS), Channel Estimation, Pilot Contamination.

I. INTRODUCTION

MASSIVE MIMO is a promising technique to achieve 5th generation targets of peak data rates up to 10Gbit/s [1], and can be defined as a system using a large number of antennas at the base station, which enables the systems capacity to serve a large number of users [2].

When comparing massive MIMO to the conventional MIMO systems, massive MIMO show several advantageous aspects. Firstly, as the number of the antennas at the base station goes to high values, the simplest coherent combiner and linear precoder turn out to be optimal. Secondly, by exploiting the features of the channel reciprocity, additional antennas increase the network capacity significantly without the need for additional feedback overhead. Thirdly, enabling the power reduction in the uplink and in the downlink can provide the potential for small-cell size shrinking [3].

The major limiting factor in massive MIMO is the availability of accurate, instantaneous channel state information (CSI) at the base station, regardless of whether the CSI is used for the uplink detection process or for the downlink precoding process. The CSI is typically acquired by transmitting predefined pilot signals and estimating the channel coefficients from the re-ceived signals by applying an appropriate estimation algorithm [1]-[3].

The channel estimation accuracy depends on having perfect orthogonal pilots allocated to the users; however, to achieve a high spectral efficiency, the same carrier frequency is to be used in the neighbouring cells by following a specific reuse pattern. This leads to the creation of a spatially correlated inter-cell interference, known as pilot contamination, which reduces the estimation performance and spectral efficiency [1]-[3].

The pilot contamination problem was analyzed in [4] and it has shown that the precoding downlink signal of the base station in the serving cell contaminated the received signal of the users roaming in other cells. The authors of [5] analyzed the pilot contamination problems in multi-cell massive MIMO systems relying on a large antennas at the base station, and demonstrated that the pilot contamination problem persisted in large-scale MIMO [6].

Compressed sensing (CS) is an advanced theory that has important applications in many areas of engineering. Using CS, sparse or compressible signals can be recovered from incoherent measurements with far fewer samples than possible via the conventional Nyquist rate. CS has found its charm in the communications field, and the premise here is that CS allows for accurate system parameter estimation with less training, resulting in addressing the pilot contamination problem and improving the bandwidth efficiency [7]. However, the classical CS algorithms require prior knowledge of channel sparsity which is usually unknown in practical scenarios.

To make the CS-based channel estimation more practical for massive MIMO systems, in this paper, we propose an improved channel estimation scheme based on the theory of BCS (Bayesian compressed sensing) that introduces relevance vector machines (RVM) and statistical Learning Information (SLI) into standard CS, whereby a priori information regarding the channel sparsity can be exploited for more reliable channel recovery [8]-[10].

The goal of this paper is to mitigate the pilot contamination problem through practical usage of CS theory using BCS algorithm. Compared to the classical based scheme, the simulation results indicate that the proposed channel estimation method provides an improved estimation accuracy and can address the
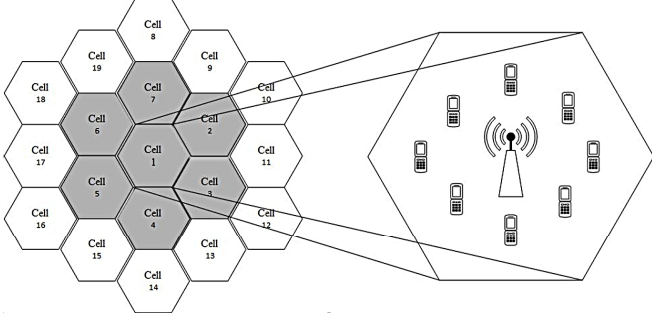
Fig. 1. Illustration of the system model of a multi-cell multi-user massive MIMO that consists of 19 cells; each cell contains eight UEs that are equally spaced on a circle with a radius of 100 m [5].

pilot contamination problem.

The remainder of this paper is organized as follows. The multi-cell massive MIMO system model is presented in Section II. The BSC-based channel estimation details are addressed in Section III, and the final conclusions are drawn in Section V.

The following notation is adopted throughout the letter: for any matrix A, $A_{i,j}$ denotes the (i,j)th element, the superscripts $(.)^H$ and $(.)^{-1}$ denote the conjugate transpose and the inverse operation, respectively,$(.)^H$ is the conjugate transpose and I denotes a diagonal matrix. The Frobenius and spectral norms of a matrix x are denoted by $\|x\|_F$ and $\|x\|_2$ respectively.$E\{.\}$ has been employed to denote expectation with regard to all random variables within the brackets. A Gaussian stochastic variable $\mathbf{o}$ is the denoted by $\mathbf{o} \sim CN(r,q)$, where r is the mean and q is the variance. Also, a random vector x having the prober complex Gaussian distribution of mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ is indicated by $x \sim N(x; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, where, $N(x; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{det(\pi\boldsymbol{\Sigma})}e^{-(x-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(x-\boldsymbol{\mu})}$, for simplicity we refer to $N(x; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ as $x \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

II. MASSIVE MIMO SYSTEM MODEL

Following the system model of [11] and [6] with modification, we consider a multi-cell massive MIMO-OFDM system with C cells. Each cell is comprised of M antennas at the base station and N single antenna users spaced equally on a circle with a radius of 100 m. The system applies a time division duplex (TDD) mode to exploit channel reciprocity, as shown in Fig.1. The uplink channel is used for the pilot-based channel estimation, and the received signal at the base station in cell c^* can be mathematically modelled as

$$y_{c^*} = x_{c^*} F h_{c^*,c^*} + \sum_{c=1,c\neq c^*}^{C} x_c F h_{c^*,c} + n \quad (1)$$

where $x_{c^*} \in C^{K \times M}$ consists of the pilot signal that is used for channel estimation sent by the users to the base station at

the cell under study c^*, $x_c \in C^{K \times M}$ consists of the pilot signal received at the base station at c^* from the other C cells, K is the number of the subcarrier, $F \in C^{L \times K}$ represents the matrix consisting the Discrete Fourier Transform (DFT) matrix and can easily given by $\frac{1}{\sqrt{K}} * [0, e^{-j2\pi*k/K}, ..., e^{-j2\pi*k(L-1)/K}]$, the L length of the number of the path, the term h_{c^*,c^*} is the channel impulse response (CIR) between users and the base station at cell under study c^* given by $h_{c^*,c^*} = [h_{c^*,c^*}(1), h_{c^*,c^*}(2)...h_{c^*,c^*}(L)]$, while the term $h_{c^*,c}$ is the (CIR) between users from the other C cells and base station at the c^*. The channel coefficient is modelled as $h_{c^*,c^*} = \sqrt{\beta_{c^*,c^*} g_{c^*,c^*}}$, where β_{c^*,c^*} model the path-loss and shadowing (large-scale fading) that are assumed to be known at the base station, while the term g_{c^*,c^*} are assumed to be independently identically distribution(i.i.d) (small-scale fading). The term $n \in C^{K \times 1}$ is an ergodic process that consists of independent receiver noise $n \sim CN(0, \sigma_{noise}^2 I)$. We defined the variance interference as σ_{interf}^2 caused during pilot transmission, we define the overall covariance matrix given as $\sigma = \sigma_{noise}^2 + \sigma_{interf}^2$, and the average power is $P^{UE} = \{|x_{c^*}|^2\}$.

III. CHANNEL ESTIMATION BASED ON BCS VIA AN RVM ALGORITHM

In this section, an RVM algorithm is presented in the context of massive MIMO channel estimation. The channel response in massive MIMO can be considered to be sparse, since it depends on a few dominant taps. Based on Bayes' rule, the full posterior distribution over unknown areas of interest for the problem at hand can be given as

$$P(h|y, \beta, \sigma^2) = \frac{P(y|h, \sigma^2)P(h|\beta)}{P(y|\beta, \sigma^2)} \sim N(\mu, \Sigma) \quad (2)$$

where β represents the hyperparameters that control the inverse variance of each channel coefficient, where the mean μ and the covariance Σ are given by

$$\mu = \beta\Sigma x_p f y \quad (3)$$

$$\Sigma = (A + \beta x_p{}^H x_p f^H f)^{-1} \quad (4)$$

where $A = diag(\beta_0, \beta_1, \beta_2, ..., \beta_N)$.

Based on the assumption of the RVM approach in [12]-[14] the term $P(h|\beta)$ follows zero- mean Gaussian distribution and can be expressed as

$$P(h|\beta) = \prod_{i=1}^{N}(h_i|0, \beta_i^{-1}) = (2\pi)^{\frac{-N}{2}} \prod_{i=1}^{N} \beta_i^{\frac{1}{2}} exp[\frac{-1}{2}h^T\beta_i h] \quad (5)$$

$$= (2\pi)^{\frac{-N}{2}} \prod_{i=1}^{N} \beta_i^{\frac{1}{2}} exp[\frac{-1}{2}h^T\beta_i h] \quad (6)$$

Based on the structured model of BCS in [12], β follows gamma distribution Γ with the shape parameter α and scale parameter a and b, that is

$$P(\boldsymbol{h}|a,b) = \prod_{i=1}^{N} N(\omega_l|0, \alpha_l^{-1}), \tag{7}$$

and the conditional distribution of $\boldsymbol{h}$ can be expressed as

$$P(\boldsymbol{h}|a,b) = \prod_{i=1}^{N} \int_0^\infty N(\boldsymbol{h_i}|0, \alpha_i^{-1}).\Gamma(\alpha_i|a,b)d\alpha_i \tag{8}$$

where $N(\boldsymbol{h_i}|0, \alpha_i^{-1})$ represents the likelihood function

To compute the full distribution approximately, the term of the marginal likelihood $P(\boldsymbol{h}|\beta,\sigma^2)$ needs to be approximated using a type$-II$ maximum likelihood procedure by integrating over the channel coefficients $\boldsymbol{h}$ as

$$P(\boldsymbol{h}|\beta,\sigma^2) = \int_{-\infty}^{\infty} P(\boldsymbol{y}|\boldsymbol{h},\sigma^2)P(\boldsymbol{h}|\beta)dh \tag{9}$$

According to the probability theory, the Gaussian likelihood function of $\boldsymbol{y}$ can be written as

$$P(\boldsymbol{y}|\boldsymbol{h},\sigma^2) = \left(\frac{2\pi}{\beta}\right)^{\frac{-N}{2}} exp\left(\frac{-\beta}{2}||\boldsymbol{y}-\boldsymbol{x_p}\boldsymbol{f}\boldsymbol{h}||_2^2\right) \tag{10}$$

By substituting (6) and (10) into (9), marginal likelihood $P(\boldsymbol{h}|\beta,\sigma^2)$ can be expressed as

$$P(\boldsymbol{h}|\beta,\sigma^2) = \left(\frac{\beta}{2\pi}\right)^{\frac{N}{2}} \left(\frac{1}{2\pi}\right)^{\frac{N}{2}} \prod_{i=1}^{N} \beta_i^{\frac{1}{2}} \int_{-\infty}^{\infty} exp\left(\frac{-\beta}{2}||\boldsymbol{y}-\boldsymbol{x_p}\boldsymbol{f}\boldsymbol{h}||_2^2\right)$$
$$+\frac{1}{2}\boldsymbol{h^T}\beta_i\boldsymbol{h}) \tag{11}$$

assuming

$$E(\boldsymbol{h}) = \frac{\beta}{2}||\boldsymbol{y}-\boldsymbol{x_p}\boldsymbol{f}\boldsymbol{h}||_2^2) + \frac{1}{2}\boldsymbol{h^H}\beta_i\boldsymbol{h} \tag{12}$$

$$E(\boldsymbol{h}) = \frac{1}{2}(\beta\boldsymbol{y^H}\boldsymbol{y}-2\beta\boldsymbol{y^H}\boldsymbol{x_p}\boldsymbol{f}\boldsymbol{h}+\beta\boldsymbol{x_p^H}\boldsymbol{f^H}\boldsymbol{h^H}\boldsymbol{x_p}\boldsymbol{f}\boldsymbol{h}+\boldsymbol{h^H}\beta\boldsymbol{h})$$

using $\Sigma = (A + \beta\boldsymbol{x_p^H}\boldsymbol{x_p}\boldsymbol{f^H}\boldsymbol{f})^{-1}$ and $I = \Sigma^{-1}\Sigma$

$$E(\boldsymbol{h}) = \frac{1}{2}(\beta\boldsymbol{y^H}\boldsymbol{y} - 2\beta\boldsymbol{y^H}\boldsymbol{x_p}\boldsymbol{f}\boldsymbol{h}\Sigma^{-1}\Sigma + \boldsymbol{h^H}\boldsymbol{h}\Sigma^{-1}) \tag{13}$$

substituting (3) into (14) yield

$$E(\boldsymbol{h}) = \frac{1}{2}(\beta\boldsymbol{y^H}\boldsymbol{y}-\mu^H\mu\Sigma^{-1})+\frac{1}{2}(\boldsymbol{h}-\mu)^H\Sigma^{-1}(\boldsymbol{h}-\mu) \tag{14}$$

the integration can then be given as

$$\int_{-\infty}^{\infty} exp(-E(\boldsymbol{h}))d\boldsymbol{h} = exp[-(\frac{1}{2}(\beta\boldsymbol{y^H}\boldsymbol{y} - \mu^H\mu\Sigma^{-1})]$$
$$* (2\pi)^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}} \tag{15}$$

Thus, the marginal likelihood can be given as

$$P(\boldsymbol{h}|\beta,\sigma^2) = \left(\frac{\beta}{2\pi}\right)^{\frac{N}{2}} \left(\frac{1}{2\pi}\right)^{\frac{N}{2}} \prod_{i=1}^{N} \beta_i^{\frac{1}{2}} \int_{-\infty}^{\infty} exp(\beta\boldsymbol{y^H}\boldsymbol{y} - \mu^H\mu\Sigma^{-1})$$
$$* (2\pi)^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}} \tag{16}$$

Then, μ and Σ (which are the mean and the variance estimation of the original channel coefficient, respectively.To get the estimation of the original signal, β and σ^2 should be estimated iteratively until the convergence criteria has been satisfied.

Algorithm 1: Relevance Vector Machine

INPUTS:
1) Pilot Signal $\boldsymbol{x_p}$
2) Observation Matrix $\boldsymbol{\phi} = \boldsymbol{f}$
OUTPUTS: σ, β
Initial Configuration:
1: Select a suitable value for convergence $\boldsymbol{\delta}$.
2: Select a start value for $\boldsymbol{\sigma^2}$ and $\boldsymbol{\beta}$
3: **repeat**
4: Update $\beta_i = \frac{I - \beta_i\Sigma_i i}{\mu_i^2}$ and $\sigma_i^2 = \frac{(N-I+\Sigma_i A_i)}{||\boldsymbol{y}-\boldsymbol{x_p}\boldsymbol{f}\boldsymbol{h}||_2^2}$
5: **Until** $\delta < \Sigma_{i=1}\sigma_i^{2n+1} - \sigma_i^{2n}$
OUTPUTS: Obtain β_i and σ_i^2

Algorithm 2: Bayesian Channel Estimator

INPUTS: $\boldsymbol{x_c}, \boldsymbol{y_c^*}$
Initial Configuration:
1: Select a suitable value for RVM as step 1-2 in Algorithm 1.
3: **repeat**
4: **RVM Algorithm:** Generate $P(\boldsymbol{h}|\beta,\sigma^2)$ using steps 4-6 in Algorithm 1.
5: **Until** $\delta < \Sigma_{i=1}\sigma_i^{2n+1} - \sigma_i^{2n}$
6: Compute $\mu = \beta\Sigma\boldsymbol{x_p}\boldsymbol{f}\boldsymbol{y}$
7: Compute $\Sigma = (A + \beta\boldsymbol{x_p}^H\boldsymbol{x_p}\boldsymbol{f}^H\boldsymbol{f})^{-1}$
8: **Bayes' Rule:** Generate $P(\boldsymbol{h}|\boldsymbol{y},\beta,\sigma^2) \sim N(\mu,\Sigma)$
9: Compute $\hat{\boldsymbol{h}} = E(P(\boldsymbol{h}|\boldsymbol{y},\beta,\sigma^2))$
OUTPUTS: Return the Estimated Channel $\hat{\boldsymbol{h}}$

β can be obtained by differentiating the log marginal likelihood with regard to β and α, and equating it to zero as follows

$$\frac{\partial}{\partial\beta_i} \ln P(\boldsymbol{y}|\beta,\sigma^2) = \frac{1}{2\beta_i} - \frac{1}{2}\Sigma_i - \frac{1}{2}m_i^2 = 0$$

$$\beta_i = \frac{I - \beta_i\Sigma_i i}{\mu_i^2} \tag{17}$$

While $\sigma.^2$ is obtained by differentiate (17) with regard to β and set these derivations to zero

$$\frac{\partial}{\partial \beta_i} \ln P(\boldsymbol{y}|\beta,\sigma^2) = \quad \frac{1}{2}\left(\frac{N}{\beta} - ||\boldsymbol{y} - \boldsymbol{x_p}\boldsymbol{f}\boldsymbol{h}||_2^2\right)$$
$$-tr(\Sigma\boldsymbol{x_p^H}\boldsymbol{f^H}\boldsymbol{h^H}\boldsymbol{x_p}\boldsymbol{f}\boldsymbol{h}) = 0$$

The argument of the $tr(.)$ can be simplified

$$\Sigma\boldsymbol{x_p^H}\boldsymbol{f^H}\boldsymbol{h^H}\boldsymbol{x_p}\boldsymbol{f}\boldsymbol{h} = \Sigma\boldsymbol{x_p^H}\boldsymbol{f^H}\boldsymbol{h^H}\boldsymbol{x_p}\boldsymbol{f}\boldsymbol{h} + \beta^{-1}\Sigma A$$
$$-\beta^{-1}\Sigma A = (I - \Sigma A)\beta^{-1}$$

then,

$$\sigma_i^2 = \frac{(N - I + \Sigma_i A_i)}{||\boldsymbol{y} - \boldsymbol{x_p}\boldsymbol{f}\boldsymbol{h}||_2^2} \tag{18}$$

Then the estimated channel based on Bayesian estimation approaches to minimize the mean square error (MSE) can be given as [15]

$$\hat{\boldsymbol{h}} = E(P(\boldsymbol{h}|\boldsymbol{y},\beta,\sigma^2)) \tag{19}$$

Further details of the BCS algorithm can be found in [12]-[14].

IV. NUMERICAL RESULTS

To verify the accuracy of our analytical results, the simulation parameters can be summarized as follows: the number of antennas is 100, the number of users is 10, the number of the paths is 10 and the number of subcarrier K is 100. The simulation results are obtained by averaging over 1000 realizations.
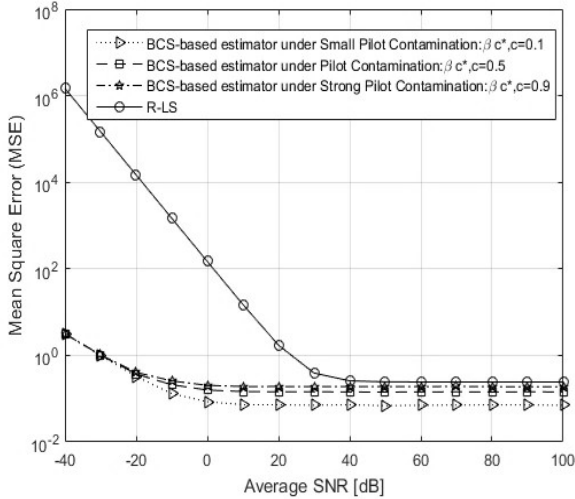

Fig. 2. MSE performance comparison between BSC and LS versus SNR.

To compare the accuracy of the channel estimation techniques, the normalized (MSE) is used for performance evaluation and is computed as
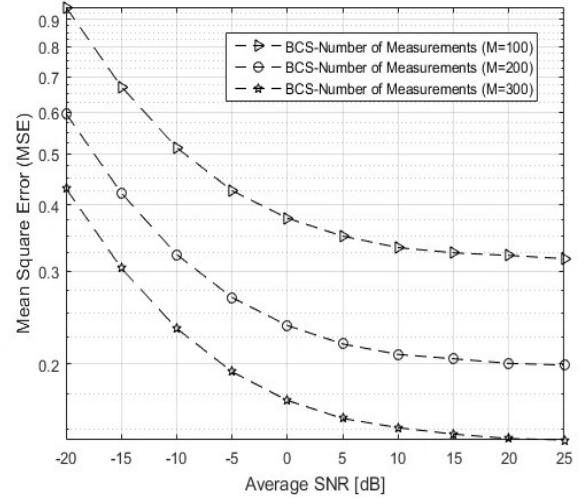

Fig. 3. MSE of BSC for different values of the number of antennas at the base station versus SNR.
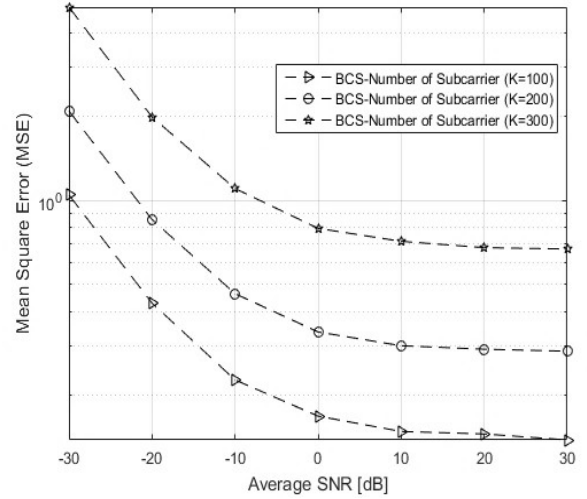

Fig. 4. MSE of BSC for different values of the number of subcarrier

$$MSE = \frac{1}{M*N}\sum_{i=1}^{M*N}\frac{||\hat{\boldsymbol{h}} - \boldsymbol{h}||_2^2}{||\boldsymbol{h}||} \tag{20}$$

Fig. 2 shows the MSE performance comparison between a BCS-based channel estimation of three scenarios of small pilot contamination ($\boldsymbol{\beta_{c*,c*}} = 1$ and $\boldsymbol{\beta_{c*,c}} = 0.1$), strong pilot contamination ($\boldsymbol{\beta_{c*,c*}} = 1$ and $\boldsymbol{\beta_{c*,c}} = 0.5$), very strong pilot contamination ($\boldsymbol{\beta_{c*,c*}} = 1$ and $\boldsymbol{\beta_{c*,c}} = 0.9$) and a regularized least square (RLS)-based estimator with no pilot contamination as a benchmark. The results have shown significant improvement of estimation accuracy and addressing the pilot contamination problem for SNR values of -40dB to 100 dB, furthermore, the results still shown enhanced estimation performance for high SNR.

Fig. 3 demonstrates the MSE of the BSC-based channel estimation versus SNR for three scenario of different setting to the number of antennas at the base station (which is the number of measurements of CS) of 100, 150 and 200, ($\beta_{c^*,c^*} = 1$ and $\beta_{c^*,c} = 0.1$), it can be seen that the estimation accuracy of the proposed algorithm is enhanced by increasing the number of the number of measurements.

Fig. 4. shows the (MSE) performance versus SNR with different value of setting to the number of subcarrier (K=100,200 and 300), ($\beta_{c^*,c^*} = 1$ and $\beta_{c^*,c} = 0.1$), the results prove that the estimation accuracy is performed better with low values of the number of subcarrier.

V. CONCLUSION

In this paper, we proposed a BCS-based channel estimation algorithm for multi-cell multi-user massive MIMO. The simulation results has revealed that the BCS-based channel estimation algorithm has tremendous improvement over the conventional-based channel estimation algorithms and can address the pilot contamination problem. In addition, the number of measurements should be selected wisely to achieve the optimum estimation accuracy.

REFERENCES

[1] Nokia Solutions and Networks, "Technology Vision 2020 White paper", pp. 116, 2013.

[2] V. Jungnickel, K. Manolakis, W. Zirwas, B. Panzner, V. Braun, M. Lossow, M. Sternad, R. Apelfr ojd, and T. Svensson, The role of small cells, coordinated multipoint, and massive MIMO in 5G,IEEE Communications Magazine, vol. 52, no. 5, pp. 4451, May 2014.

[3] H. Zhang, S. Gao, D. Li, H. Chen and L. Yang, "On Superimposed Pilot for Channel Estimation in Multi-cell Multiuser MIMO Uplink: Large System Analysis", IEEE Transaction on Vehicular Technology, pp. 99, 2015.

[4] Jose, J., Ashikhmin, A., Marzetta, T.L., et al.: "Pilot contamination problem in multi-cell TDD systems", Proc. IEEE Int. Symp. Inf. Theory, Seoul, pp. 2184 -2188, June 2009.

[5] J. Jose, A. Ashikhmin, T. L. Marzetta., et al.: "Pilot contamination and precoding in multi-cell TDD systems, IEEE Trans. Wireless Communication ", pp. 2640 -2651,2011.

[6] J. Zhang, B. Zhang, S. Chen, et al.: Pilot contamination elimination for large-scale multiple-antenna aided OFDM systems, IEEE J. Sel. Topics Signal Processing", pp. 759 -772,2014.

[7] Nguyen, Sinh, Compressive Sensing for Multi-channel and Large-scale MIMO Networks, PhD. thesis, Dept. of Elect. And Comp. Eng., Concordia Univ., Montreal, Quebec, Canada, 13 Jan 2014.

[8] Zhenkai Fan, Zhaohua Lu and Yanjun Han, "Accurate channel estimation based on Bayesian Compressive Sensing for next-generation wireless broadcasting systems", 2014 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), pp. 1-5, Beijing, China, 25-27 June 2014.

[9] Zhenkai Fan, Zhaohua Lu and Yanjun Han, "Reliable channel estimation based on Bayesian compressive sensing for TDS-OFDM systems", 2014 IEEE International Conference on Communication Systems (ICCS), pp. 620 - 624, Macau, 19-21 Nov. 2014.

[10] Xiantao Cheng; Mengyao Wang and Yong Liang Guan, "Ultrawideband Channel Estimation: A Bayesian Compressive Sensing Strategy Based on Statistical Sparsity, IEEE Trans. Veh. Technol., vol. 64, no. 5, pp. 1819 - 1832, 12 May 2015.

[11] Emil Bjrnson, Jakob Hoydis, Marios Kountouris, Mrouane Debbah, "Massive MIMO Systems with Non-Ideal Hardware: Energy Efficiency, Estimation, and Capacity Limits", IEEE Transactions on Information Theory, vol. 60, no. 11, pp. 7112-7139, 2014.

[12] S. Ji , Y. Xue and L. Carin , "Bayesian compressive sensing", IEEE Trans. Signal Process. , vol. 56 , no. 6 , pp.2346 -2356 , 2008

[13] Tristan Fletcher, "Relevance Vector Machines Explained", October 19, 2010, [online]. Avialable: http://www.tristanfletcher.co.uk/RVM

[14] An introduction to Statistical Learning with Applications in R, G. Casella S. Fienberg I. Olkin, Springer, 2013.

[15] S. Kay, 'Fundamentals of Statistical Signal Processing: Estimation Theory', Englewood Cliffs, NJ: Prentice Hall, 1993.