Channel Estimation for mmWave MIMO with Transmitter Hardware Impairments

Yue Wu, Yuantao Gu, and Zhaocheng Wang

***Abstract*—This paper considers the problem of channel estimation for millimeter wave (mmWave) multiple input multiple output (MIMO) systems under a transmitter impairments model. Specifically, taking the transmitter hardware impairments into account, the performance of conventional pilots-based channel estimation scheme will be degraded due to the destroyed training pilots. By exploiting the sparsity of mmWave channel in angular domain, a new channel estimation algorithm based on the Bayesian compressive sensing (BCS) and least square estimation (LSE) is proposed. First, the expectation maximization (EM) algorithm is presented to solve the BCS problem, and the refined measurement matrix and the support of the channel vector are obtained. Next, the channel gain coefficients are estimated by using the LSE. Simulation results show that the proposed algorithm can achieve better performance compared to the conventional BCS and orthogonal matching pursuit algorithm (OMP).**

***Index Terms*—Millimeter wave, channel estimation, hardware impairment, compressed sensing.**

I. Introduction

Millimeter wave (mmWave) communication has been considered as a promising technique for the fifth generation (5G) mobile networks [1]. However, the severe signal attenuation is a critical challenge for exploiting the mmWave frequency band. Highly directional beamforming based on massive antenna integrated in transceiver is a usual solution for the path loss compensation [4], which requires accurate estimation of channel state information (CSI).

Channel estimation in mmWave systems is challenging due to the use of large antenna arrays. Recent research measurements show that mmWave channels typically appear to be sparse in the angular domain [2]. Thus, the mmWave channel estimation can be formulated as a sparse signal recovery problem, and the compressed sensing tools can be exploited to significantly reduce the estimation training overhead [4].

In order to reduce the hardware costs and energy consumption, it is more attractive to deploy large antenna arrays with inexpensive and power-efficient components. However, the use of such cheap components is more likely to induce hardware impairments, such as phase noise, nonlinear power amplifier, I/Q imbalance, quantization errors, and power losses [5]–[7]. The impact of some of these impairments can be mitigated by pre-distortion or compensation algorithms partially, but there always remains a certain amount of distortion that cannot be fully parameterized and estimated [5]. Finally, the residual impairments will destroy the training pilots and degrade the performance of channel estimation. However, there are few researches to consider it in mmWave systems. In [4], the channel estimation scheme was proposed for mmWave massive MIMO systems with ideal transceiver hardware. The authors in [5] consider the problem of channel estimation for massive MIMO systems with transceiver hardware impairments, which did not specialize for mmWave systems.

In this paper, we address the mmWave channel estimation problem with transmitter hardware impairments, and formulate the channel estimation as a sparse signal recovery problem with measurement uncertainty. A new algorithm based on Bayesian compressive sensing (BCS) and least square estimation (LSE) is proposed to estimate the mmWave channel. The proposed algorithm first solves a BCS problem, and obtains a refined measurement matrix and the support of channel vector, and next estimates the channel vector with LSE. Compared to the conventional BCS ignoring hardware impairments and OMP algorithm, the proposed algorithm provides a better performance because it considers the perturbation in measurement matrix. Simulation results verify the good performance of the proposed algorithm. To the best of our knowledge, this is the first work to investigate the channel estimation with hardware impairments in mmWave MIMO systems.

II. System Model

Consider a point-to-point mmWave MIMO communication system with N_r receive antennas and N_t transmit antennas. We assume that hybrid analog-digital architectures [7] are employed and $M_t \leq N_t$, $M_r \leq N_r$ RF chains are implemented at the transceivers. Before transmission, the transmitter linearly processes the vector of data streams $\mathbf{s} \in \mathbb{C}^{N_s \times 1}$ with a hybrid precoder $\mathbf{F} = \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}$, where $\mathbf{F}_{\mathrm{BB}} \in \mathbb{C}^{M_t \times N_s}$ denotes the digital baseband precoder and the RF precoder is presented as $\mathbf{F}_{\mathrm{RF}} \in \mathbb{C}^{N_t \times M_t}$. The discrete-time transmitted signal is

$$\mathbf{x} = \mathbf{F}\mathbf{s}, \tag{1}$$

where $\mathbf{x} \in \mathbb{C}^{N_t \times 1}$ is the training pilot signal or the communication data signal. In both cases, the covariance matrix is denoted by $\mathbf{Q} = \mathbb{E}\{\mathbf{x}\mathbf{x}^{\mathrm{H}}\}$.

A narrowband block-fading channel model is adopted in this paper, and the receiver observes the signal

$$\mathbf{r} = \mathbf{H}\mathbf{F}\mathbf{s} + \mathbf{n}, \tag{2}$$

The authors are with the Department of Electronic Engineering and Tsinghua National Laboratory for Information Science and Technology (TNList), Tsinghua University, Beijing 100084, China. This work was partially supported by National Natural Science Foundation of China (NSFC 61531166005, 61371137, 61571263, 51459003). The corresponding author of this work is Y. Gu (E-mail: gyt@tsinghua.edu.cn).

where $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ denotes the mmWave MIMO channel matrix, and $\mathbf{n}$ represents the complex additive white Gaussian noise following the distribution $\mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I}_{N_r})$. The receiver adopts a hybrid combiner $\mathbf{W} = \mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}}$ to process the received signal $\mathbf{r}$, where $\mathbf{W}_{\mathrm{BB}} \in \mathbb{C}^{M_r \times N_s}$ is a baseband combiner and $\mathbf{W}_{\mathrm{RF}} \in \mathbb{C}^{N_r \times M_r}$ is a RF combiner. Finally, the combined signal is

$$\mathbf{y} = \mathbf{W}^{\mathrm{H}}\mathbf{H}\mathbf{F}\mathbf{s} + \mathbf{W}^{\mathrm{H}}\mathbf{n}. \tag{3}$$

A. Transmit Hardware Impairments

The aggregate residual transmitter hardware impairments can be modeled as additive distortion noise, which has been used and verified by experiments in many previous works [5], [6], [8]. Based on the transceiver model (3), the actual combined signal at the receiver is

$$\mathbf{y} = \mathbf{W}^{\mathrm{H}}\mathbf{H}(\mathbf{F}\mathbf{s} + \mathbf{e}) + \mathbf{W}^{\mathrm{H}}\mathbf{n}, \tag{4}$$

where $\mathbf{e}$ denotes the additive distortion noise, which is assumed independent of the transmitted signal $\mathbf{x}$. Furthermore, measurements results [8] show that the transmit distortion noise can be modeled as Gaussian distributed well, with the key property that its power is proportional to the signal power at each antenna [6], [8]. Thus we have $\mathbf{e} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Upsilon})$, where $\mathbf{\Upsilon} \triangleq \kappa_e \mathrm{diag}(q_{11}, \ldots, q_{N_t N_t})$ with q_{ii} denotes the ith diagonal element of covariance matrix $\mathbf{Q}$, and κ_e is the proportionality coefficient. The proportionality coefficient κ_e characterizes the levels of impairments. In the long term evolution (LTE) standard, the κ_e requirements are in the range $[0.08^2, 0.175^2]$ [5]. In practical mmWave MIMO systems, since such cheap equipments are encouraged to be used, the larger values of κ_e are of interest in this paper.

B. Millimeter Wave Channel Model

In this paper, a geometric channel model with L distinct paths is adopted [3],

$$\mathbf{H} = \sum\nolimits_{l=1}^{L} \beta_l \mathbf{a}_r(\theta_l)\mathbf{a}_t^{\mathrm{H}}(\phi_l), \tag{5}$$

where β_l, ϕ_l, θ_l are the complex path gain, azimuth angle of departure, azimuth angle of arrival of the lth path, respectively. $\mathbf{a}_t(\phi_l)$ and $\mathbf{a}_r(\theta_l)$ are the antenna array response vectors at transceiver which are dependent on the array structure. For arrival angle θ_l, the uniform linear antenna array response is $\mathbf{a}_r(\theta_l) = \frac{1}{\sqrt{N_r}}[1, e^{\mathrm{j}\frac{2\pi}{\lambda}d\sin(\theta_l)}, \ldots, e^{\mathrm{j}(N_r-1)\frac{2\pi}{\lambda}d\sin(\theta_l)}]^{\mathrm{T}}$, where λ is the carrier wavelength, and d is the antenna spacing. Similarly, we can get the expression of $\mathbf{a}_t(\phi_l)$.

Next, the virtual channel representation is as well adopted that maps the time-domain channel $\mathbf{H}$ into the angle-domain $\mathbf{H}_v$ as [3]

$$\mathbf{H} = \mathbf{D}_r \mathbf{H}_v \mathbf{D}_t^{\mathrm{H}}, \tag{6}$$

where $\mathbf{D}_r \in \mathbb{C}^{N_r \times N_r}$ and $\mathbf{D}_t \in \mathbb{C}^{N_t \times N_t}$ are the DFT matrices which uniformly quantize the angular domain with the spacings of $2\pi/N_r$ at the receiver and $2\pi/N_t$ at the transmitter, respectively. In this paper, we only consider the case that the quantized $2\pi\sin(\phi_l)$ and $2\pi\sin(\theta_l)$ have the same resolutions as $\mathbf{D}_r$ and $\mathbf{D}_t$, thus the number of non-zero entries of $\mathbf{H}_v$ is L ($L \ll N_t N_r$).

C. Formulation of Channel Estimation

In order to exploit the sparse nature of the channel, we recast the RHS of (4) as

$$\begin{aligned}\mathbf{y} &\overset{(a)}{=} \left((\mathbf{F}\mathbf{s} + \mathbf{e})^{\mathrm{T}} \otimes \mathbf{W}^{\mathrm{H}}\right)\mathrm{vec}(\mathbf{H}) + \mathbf{W}^{\mathrm{H}}\mathbf{n} \\ &\overset{(b)}{=} \left((\mathbf{F}\mathbf{s} + \mathbf{e})^{\mathrm{T}} \otimes \mathbf{W}^{\mathrm{H}}\right)(\mathbf{D}_t^* \otimes \mathbf{D}_r)\,\mathbf{h} + \mathbf{W}^{\mathrm{H}}\mathbf{n},\end{aligned}$$

where $\mathbf{h} = \mathrm{vec}(\mathbf{H}_v) \in \mathbb{C}^{N_r N_t \times 1}$ is a L-sparse vector, (a) and (b) follow the identity $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^{\mathrm{T}} \otimes \mathbf{A})\mathrm{vec}(\mathbf{B})$. We assume that M symbol times are dedicated for estimating during each channel block, $\mathbf{W}^{(m)}$ and $\mathbf{x}^{(m)}$ are the combiner and the training pilot for the mth training time instant at receiver and transmitter, respectively. Stacking the M measurements into a vector, we have

$$\underbrace{\begin{bmatrix}\mathbf{y}^{(1)} \\ \vdots \\ \mathbf{y}^{(M)}\end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix}(\mathbf{x}^{(1)} + \mathbf{e}^{(1)})^{\mathrm{T}} \otimes \mathbf{W}^{(1)} \\ \vdots \\ (\mathbf{x}^{(M)} + \mathbf{e}^{(M)})^{\mathrm{T}} \otimes \mathbf{W}^{(M)}\end{bmatrix}}_{\mathbf{\Phi}} \mathbf{D}\,\mathbf{h} + \underbrace{\begin{bmatrix}\tilde{\mathbf{n}}^{(1)} \\ \vdots \\ \tilde{\mathbf{n}}^{(M)}\end{bmatrix}}_{\tilde{\mathbf{n}}}, \tag{7}$$

where $\mathbf{D} \triangleq \mathbf{D}_t^* \otimes \mathbf{D}_r$, $\tilde{\mathbf{n}}^{(m)} \triangleq (\mathbf{W}^{(m)})^{\mathrm{H}}\mathbf{n}^{(m)}$. Since the distortion noise $\mathbf{e}^{(m)}$ in the measurement matrix $\mathbf{\Phi}$ is unknown, the mmWave channel estimation problem is simplified to estimate the L-sparse vector $\mathbf{h}$ with measurement matrix uncertainty.

III. Sparse Channel Estimation

In this section, a novel algorithm is proposed to solve the channel estimation problem with transmitter hardware impairments, which is named BCS-LSE. The new algorithm first estimates the support of $\mathbf{h}$ and gets the refined measurement matrix $\mathbf{\Phi}$, and next estimates the channel gain β_l. Before listing the detail of BCS-LSE algorithm, the hierarchical Bayesian model is introduced below.

A. Sparse Bayesian Formulation

After linear transformation, the processed channel noise $\tilde{\mathbf{n}}$ is still complex Gaussian

$$p(\tilde{\mathbf{n}}|\alpha_0) = \mathcal{CN}\left(\tilde{\mathbf{n}}|\mathbf{0}, \alpha_0^{-1}\mathbf{\Sigma}_n\right), \tag{8}$$

where $\alpha_0 = \sigma_n^{-2}$, $\mathcal{CN}(\mathbf{u}|\boldsymbol{\mu}, \mathbf{\Sigma})$ denotes the probability density function (PDF) of a (circular symmetric) complex Gaussian distributed random variable $\mathbf{u} \sim \mathcal{CN}(\boldsymbol{\mu}, \mathbf{\Sigma})$ with mean $\boldsymbol{\mu}$ and covariance $\mathbf{\Sigma}$, and $\mathbf{\Sigma}_n$ is a block diagonal matrix

$$\mathbf{\Sigma}_n = \begin{bmatrix}(\mathbf{W}^{(1)})^{\mathrm{H}}\mathbf{W}^{(1)} & \ldots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \ldots & (\mathbf{W}^{(M)})^{\mathrm{H}}\mathbf{W}^{(M)}\end{bmatrix}. \tag{9}$$

Given the channel vector, channel noise variance, and the distortion noise, the distribution of the combined signal at the receiver is

$$p\big(\mathbf{y}|\mathbf{h}, \alpha_0, \mathbf{E}, \{\mathbf{x}^{(m)}, \mathbf{W}^{(m)}\}_{m=1}^{M}\big) = \mathcal{CN}\left(\mathbf{y}|\mathbf{\Phi}\mathbf{h}, \alpha_0^{-1}\mathbf{\Sigma}_n\right), \tag{10}$$

where $\mathbf{E} \triangleq [\mathbf{e}^{(1)}, \ldots, \mathbf{e}^{(M)}]$. In this paper, we assume that α_0 is unknown, with a Gamma prior placed on it as follows

$$p(\alpha_0; c, d) = \Gamma(\alpha_0|c, d), \tag{11}$$

Algorithm 1 Proposed BCS-LSE algorithm

Input: Combined signal $\mathbf{y}$, pilot signals $\mathbf{x}^{(m)}$, combiners $\mathbf{W}^{(m)}$, parameters c, d, and ρ, tolerance τ.

Output: The support set of channel vector $\hat{\mathcal{S}}$ and the channel gains $\hat{\boldsymbol{\beta}}$

1: Initialize $\mathbf{E} = \mathbf{0}$, and compute measurement matrix $\mathbf{\Phi}$ in (7), $\alpha_0 = 100/\mathrm{Var}(\mathbf{y})$, $\boldsymbol{\alpha} = |\mathbf{\Phi}^{\mathrm{H}}\mathbf{y}|/(MM_rL)$.
2: **Repeat**
3: Compute $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ in (17) and (18), respectively.
4: Compute $\boldsymbol{\alpha}^{\mathrm{new}}$, α_0^{new}, and $\mathbf{E}$ in (19), (20), and (21), respectively.
5: Refine $\mathbf{\Phi}$ with the estimated $\mathbf{E}$ in step 4.
6: **Until** $\|\boldsymbol{\alpha}^{\mathrm{new}} - \boldsymbol{\alpha}^{\mathrm{old}}\|_2/\|\boldsymbol{\alpha}^{\mathrm{old}}\|_2 < \tau$
7: Detect the support $\hat{\mathcal{S}}$ of channel vector $\mathbf{h}$ according to $\boldsymbol{\alpha}$.
8: Estimate the channel gains $\hat{\boldsymbol{\beta}} = (\mathbf{\Phi}^{\mathrm{H}}_{:,\hat{\mathcal{S}}}\mathbf{\Phi}_{:,\hat{\mathcal{S}}})^{-1}\mathbf{\Phi}^{\mathrm{H}}_{:,\hat{\mathcal{S}}}\mathbf{y}$.

where $\Gamma(\alpha_0|c,d) = [\Gamma(c)]^{-1}d^c\alpha_0^{c-1}\exp\{-d\alpha_0\}$ with $\Gamma(c) = \int_0^\infty t^{c-1}e^{-t}\mathrm{d}t$ is the Gamma function. Parameters c and d are set to small values as in [9] to make the priors non-informative.

Furthermore, in order to induce a sparse prior for the channel $\mathbf{h}$, a two-stage hierarchical prior is adopted [9]: $p(\mathbf{h};\rho) = \int p(\mathbf{h}|\boldsymbol{\alpha})p(\boldsymbol{\alpha};\rho)\mathrm{d}\boldsymbol{\alpha}$, where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{N_rN_t})$, and each h_n has a zero-mean complex Gaussian prior, with its own variance α_n

$$p(\mathbf{h}|\boldsymbol{\alpha}) = \prod\nolimits_{n=1}^{N_rN_t} \mathcal{CN}(h_n|0,\alpha_n), \tag{12}$$

and each α_n has an exponential prior with parameter ρ

$$p(\boldsymbol{\alpha},\rho) = \prod\nolimits_{n=1}^{N_rN_t} \rho\exp(-\rho\alpha_n), \rho > 0. \tag{13}$$

Such choice of prior is convenient in Bayesian inference due to its conjugacy properties [10]. Moreover, both $\Re\{\mathbf{h}\}$ and $\Im\{\mathbf{h}\}$ are Laplace priors with distributing the mass more on the axes. As a result, the channel vector $\mathbf{h}$ is sparse.

Finally, combine all the stages of the hierarchical Bayesian model, the joint distribution is

$$\begin{aligned} p(\mathbf{h},\mathbf{y},\alpha_0,\boldsymbol{\alpha},\mathbf{E}) =& p\big(\mathbf{y}|\mathbf{h},\alpha_0,\mathbf{E},\{\mathbf{x}^{(m)},\mathbf{W}^{(m)}\}_{m=1}^M\big) \\ & \cdot p(\mathbf{h}|\boldsymbol{\alpha})p(\boldsymbol{\alpha})p(\alpha_0)p(\mathbf{E}). \end{aligned} \tag{14}$$

B. BCS-LSE Channel Estimation Algorithm

In the standard BCS methods, these hyper-parameters are estimated by performing an evidence maximization procedure [10]. Specifically, by marginalizing over the channel vector $\mathbf{h}$, the logarithm of the marginal posterior probability for these hyper-parameters can be expressed as

$$\begin{aligned} &\mathcal{L}(\boldsymbol{\alpha},\alpha_0,\mathbf{E}) = \log p(\mathbf{y},\alpha_0,\boldsymbol{\alpha},\mathbf{E}) \\ &= \log\int p(\mathbf{y}|\mathbf{h},\alpha_0,\mathbf{E})p(\mathbf{h}|\boldsymbol{\alpha})p(\alpha_0)p(\boldsymbol{\alpha})p(\mathbf{E})\mathrm{d}\mathbf{h} \\ &= -\log|\mathbf{C}| - \mathbf{y}^{\mathrm{H}}\mathbf{C}^{-1}\mathbf{y} - \sum\nolimits_{m=1}^M(\mathbf{e}^{(m)})^{\mathrm{H}}\mathbf{\Upsilon}^{-1}\mathbf{e}^{(m)} \\ &\quad - \sum\nolimits_{n=1}^{N_rN_t}\rho\alpha_n - d\alpha_0 + (c-1)\log\alpha_0 + T, \end{aligned} \tag{15}$$

where $\mathbf{C} = 1/\alpha_0\mathbf{\Sigma}_n + \mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}^{\mathrm{H}}$, $\mathbf{\Lambda} = \mathrm{diag}(\boldsymbol{\alpha})$, and T is a constant independent of the hyper-parameters. The EM algorithm is exploited to maximize (15) that treats $\mathbf{h}$ as the hidden variables and maximizes $\mathbb{E}_{\mathbf{h}|\mathbf{y},\boldsymbol{\alpha},\alpha_0,\mathbf{E}}\{\log p(\mathbf{h},\mathbf{y},\alpha_0,\boldsymbol{\alpha},\mathbf{E})\}$ [9]. Note that given hyper-parameters, the posterior distribution of $\mathbf{h}$ is a complex Gaussian distribution

$$p(\mathbf{h}|\mathbf{y},\alpha_0,\boldsymbol{\alpha},\mathbf{E}) = \mathcal{CN}(\mathbf{h}|\boldsymbol{\mu},\mathbf{\Sigma}), \tag{16}$$

with

$$\boldsymbol{\mu} = \alpha_0\mathbf{\Sigma}\mathbf{\Phi}^{\mathrm{H}}\mathbf{\Sigma}_n^{-1}\mathbf{y}, \tag{17}$$

$$\mathbf{\Sigma} = \left(\alpha_0\mathbf{\Phi}^{\mathrm{H}}\mathbf{\Sigma}_n^{-1}\mathbf{\Phi} + \mathbf{\Lambda}^{-1}\right)^{-1}. \tag{18}$$

For notation simplification, we discard the subscript of the condition expectation $\mathbb{E}_{\mathbf{h}|\mathbf{y},\boldsymbol{\alpha},\alpha_0,\mathbf{E}}\{\cdot\}$ in the subsequent. The update of $\boldsymbol{\alpha}$ is as follows

$$\alpha_n^{\mathrm{new}} = \frac{\sqrt{1+4\rho\mathbb{E}\{\|h_n\|_2^2\}}-1}{2\rho}, n = 1,\ldots,N_rN_t, \tag{19}$$

where $\mathbb{E}\{\|h_n\|_2^2\} = \mu_n^2 + \sigma_{n,n}$, $\sigma_{n,n}$ denotes the nth diagonal element of $\mathbf{\Sigma}$. For $\mathbf{E}$ and α_0, their estimate maximizes $\mathbb{E}\{\log(p(\mathbf{y}|\mathbf{h},\alpha_0,\mathbf{E})p(\alpha_0)p(\mathbf{E}))\}$. A sequential update strategy which first updates α_0, and then updates $\mathbf{E}$ is adopted here. Such operation arises from the idea of expectation conditional maximization (ECM) algorithm, and it is clear that the objective function in (15) is guaranteed to increase at each iteration. The update of α_0 is given by

$$\alpha_0^{\mathrm{new}} = \frac{MM_r + c - 1}{\mathbb{E}\{\|\mathbf{y}-\mathbf{\Phi}\mathbf{h}\|_2^2\} + d}, \tag{20}$$

where $\mathbb{E}\{\|\mathbf{y}-\mathbf{\Phi}\mathbf{h}\|_2^2\} = \|\mathbf{y}-\mathbf{\Phi}\boldsymbol{\mu}\|_2^2 + \mathrm{Tr}\{\mathbf{\Phi}\mathbf{\Sigma}\mathbf{\Phi}^{\mathrm{H}}\}$. After the update of α_0, the update of $\mathbf{E}$ minimizes

$$\begin{aligned} \sum\nolimits_{m=1}^M(\mathbf{e}^{(m)})^{\mathrm{H}}\mathbf{\Upsilon}^{-1}\mathbf{e}^{(m)} + \mathbb{E}\Big\{\alpha_0^{\mathrm{new}}\sum\nolimits_{m=1}^M\|\mathbf{y}^{(m)} - \\ \big((\mathbf{x}^{(m)}+\mathbf{e}^{(m)})^{\mathrm{T}}\otimes\mathbf{W}^{(m)}\big)\mathbf{D}\mathbf{h}\|_2^2\Big\}. \end{aligned} \tag{21}$$

The objective function in (21) can be decomposed to M decouple functions, each of which is quadratic and only related with $\mathbf{e}^{(m)}$

$$\left(\mathbf{e}^{(m)}\right)^{\mathrm{H}}\mathbf{P}^{(m)}\mathbf{e}^{(m)} - 2\Re\left\{\left(\mathbf{v}^{(m)}\right)^{\mathrm{H}}\mathbf{e}^{(m)}\right\} + C^{(m)}, \tag{22}$$

where $C^{(m)}$ is a constant term independent of $\mathbf{e}^{(m)}$, $\mathbf{P}^{(m)}$ is a positive semi-definite matrix, and

$$\mathbf{P}^{(m)} = \mathbf{\Upsilon}^{-1} + \alpha_0^{\mathrm{new}}\left((\mathbf{G}^{(m)})^{\mathrm{H}}\mathbf{G}^{(m)} + \mathbf{O}^{(m)}\right), \tag{23}$$

$$\mathbf{v}^{(m)} = \alpha_0^{\mathrm{new}}\left((\mathbf{G}^{(m)})^{\mathrm{H}}\mathbf{b}^{(m)} - \mathbf{u}^{(m)}\right). \tag{24}$$

The $\mathbf{G}^{(m)}$, $\mathbf{b}^{(m)}$, $\mathbf{O}^{(m)}$, and $\mathbf{u}^{(m)}$ in (23), (24) are defined as

$$\mathbf{G}^{(m)} \triangleq [\mathbf{W}^{(m)}\mathbf{D}_1\boldsymbol{\mu},\ldots,\mathbf{W}^{(m)}\mathbf{D}_{N_t}\boldsymbol{\mu}], \tag{25}$$

$$\mathbf{b}^{(m)} \triangleq \mathbf{y}^{(m)} - ((\mathbf{b}^{(m)})^{\mathrm{T}}\otimes\mathbf{W}^{(m)})\mathbf{D}\boldsymbol{\mu}, \tag{26}$$

$$[\mathbf{O}^{(m)}]_{i,j} \triangleq \mathrm{Tr}\{\mathbf{W}^{(m)}\mathbf{D}_i\mathbf{\Sigma}(\mathbf{W}^{(m)}\mathbf{D}_j)^{\mathrm{H}}\}, \tag{27}$$

$$[\mathbf{u}^{(m)}]_i \triangleq \mathrm{Tr}\{((\mathbf{b}^{(m)})^{\mathrm{T}}\otimes\mathbf{W}^{(m)})\mathbf{D}\mathbf{\Sigma}(\mathbf{W}^{(m)}\mathbf{D}_i)^{\mathrm{H}}\}, \tag{28}$$

where $\mathbf{D}_i$ denotes the submatrix of $\mathbf{D}$ from the $((i-1)\times N_r+1)$th column to the $(i\times N_r)$th column.

A summary of the BCS-LSE algorithm is listed in

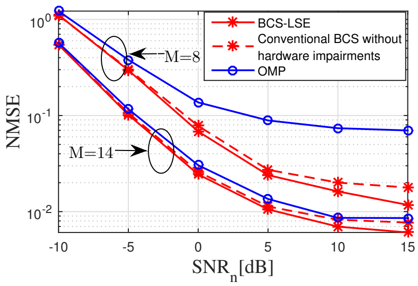

Fig. 1. Average NMSE performance comparison of BCS-LSE, conventional BCS, and OMP versus SNR_n, $\mathrm{SNR}_e = 12\mathrm{dB}$.

Alogrithm 1. In the iterations, it is observed that many of the α_n tend to zero, only a relatively small set of h_n, for which the corresponding α_n remains relatively large, and the L indices of the first largest α_n can be treated as the support of $\mathbf{h}$. Once the support of $\mathbf{h}$ has been recovered, the LSE method is adopted to estimate the values of the path gain coefficients.

IV. Numerical Results

In this section, Monte-Carlo simulations are carried out to investigate the performance of the proposed BCS-LSE channel estimation algorithm. In simulations, $N_t = 8$, $N_r = 32$, $M_r = 12$, $L = 2$, $c = d = 1 \times 10^{-4}$, $\rho = 100$, $\tau = 10^{-4}$ and $\mathrm{SNR}_n \triangleq 10\log_{10}(\sum_{m=1}^{M} \|\mathbf{H}\mathbf{b}^{(m)}\|_2^2/(MM_r\sigma_n^2))$, $\mathrm{SNR}_e \triangleq 10\log_{10}(1/\kappa_e)$. The positions and values of non-zeros elements in channel vector $\mathbf{h}$ are generated by a random way with distribution following $\mathcal{CN}(0,1)$. The pilots $\mathbf{x}^{(m)}$ are generated with distribution following $\mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_t})$, and the entries of the hybrid combiners $\mathbf{W}^{(m)}$ are generated with distribution following $\mathcal{CN}(0,1)$. The performance metric is the normalized mean square error (NMSE) of $\mathbf{h}$, given by $\mathbb{E}\{\|\mathbf{h} - \hat{\mathbf{h}}\|_2^2/\|\mathbf{h}\|_2^2\}$, and the support recovery rate for $\mathcal{S}$, i.e., $p(\mathcal{S} = \hat{\mathcal{S}})$. The conventional BCS algorithm and the OMP algorithm are adopted for comparison, and both ignore the hardware impairments.

Figure 1 shows the average channel estimation NMSE of our proposed BCS-LSE algorithm, conventional BCS algorithm, and OMP algorithm under $\mathrm{SNR}_e = 12\mathrm{dB}$. From Fig. 1, it can be observed that the proposed BCS-LSE estimator obtains better NMSE performance than the conventional BCS since it accounts for the matrix uncertainty of $\mathbf{\Phi}$. The OMP performs the worst, although it is very simple.

Figure 2 shows the performance comparison under $\mathrm{SNR}_e = 8\mathrm{dB}$. It can be observed that the proposed BCS-LSE algorithm achieves better effectiveness compared to the conventional BCS algorithm and the OMP algorithm with the presence of severe hardware impairments. Figure 3 investigates the support recovery rate comparison under $\mathrm{SNR}_e = 12\mathrm{dB}$ and $M = 8$. It can be observed that the proposed BCS-LSE algorithm obtain the highest support recovery rate.

V. Conclusion

In this paper, the hardware impairments at the transmitter is considered for mmWave MIMO channel estimation. The channel estimation has been formulated into a compressed sensing problem with measurement matrix uncertainty, and a new algorithm named BCS-LSE has been proposed to solve it, which can effectively combat the hardware impairments at the transmitter. Simulation results have validated that our proposed algorithm provides much better performance than the conventional BCS algorithm and the OMP algorithm.

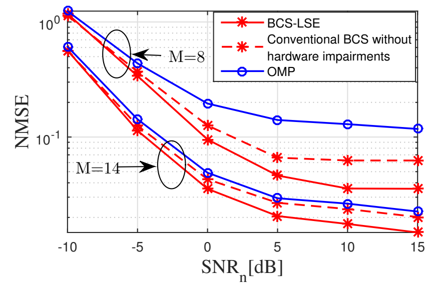

Fig. 2. Average NMSE performance comparison of BCS-LSE, conventional BCS, and OMP versus SNR_n, $\mathrm{SNR}_e = 8\mathrm{dB}$.

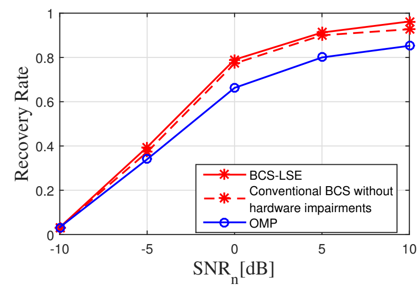

Fig. 3. Recovery rate comparison of BCS-LSE, conventional BCS, and OMP versus SNR_n, M = 8 and $\mathrm{SNR}_e = 12\mathrm{dB}$.

References

[1] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G.N. Wong, J.K. Schulz, M. Samimi and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: it will work!" *IEEE Access*, vol. 1, pp. 335-349, May. 2013.

[2] M. R. Akdeniz, Y. Liu, M.K.Samimi, S.Sun, S. Rangan, T.S. Rappaport and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164-1179, June. 2014.

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

[4] Y. Han and J. Lee, "Two-stage compressed sensing for millimeter wave channel estimation," in *Proc. IEEE ISIT*, Barcelona, Spain, July. 2016, pp. 860-864.

[5] E. Bjrnson, 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.

[6] C. Studer, M. Wenk and A. Burg, "MIMO transmission with residual transmit-RF impairments," in *Proc. IEEE WSA*, Bremen, Germany, Feb. 2010, pp. 189-196.

[7] A. Garcia-Rodriguez, V. Venkateswaran, P. Rulikowski, and C. Masouros, "Hybrid analogdigital precoding revisited under realistic RF modeling," *IEEE Wireless Commun. Lett.*, vol. 5, no. 5, pp. 528531, Oct. 2016.

[8] M. Wenk, *MIMO-OFDM Testbed: Challenges, Implementations, and Measurement Results*, ser. Series in microelectronics. Hartung-Gorre, 2010.

[9] Z. Yang, L. Xie and C. Zhang, "Off-grid direction of arrival estimation using sparse bayesian inference," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 38-43, Jan. 2013.

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