DITHERED BEAMFORMING FOR CHANNEL ESTIMATION IN MMWAVE-BASED MASSIVE MIMO

Evangelos Vlachos and John Thompson

Institute for Digital Communications, University of Edinburgh, Edinburgh, EH9 3JL, UK
{e.vlachos,j.s.thompson}@ed.ac.uk

ABSTRACT

In this work we consider the challenging problem of channel estimation at the receiver of a massive multiple-input multiple-output system with hybrid analog/digital beamforming and low-resolution quantization. We propose a dithered beamforming architecture, where random control signals are injected to the analog part of the receiver beamformer and to the analog-to-digital converters to introduce randomness into the signal capturing process and combat the stair-case quantization effects. The statistical properties of the dithered output are captured via an Expectation-Maximization approximation of the maximum a-posteriori estimator. A low-complexity algorithm is proposed which exhibits performance close to the oracle-based least-squares estimation of the sparse channel.

***Index Terms*—** low resolution analog-to-digital converter (ADC), hybrid analog/digital beamforming, millimeter wave (mmWave) massive MIMO

1. INTRODUCTION

Fifth generation (5G) networks are expected to integrate almost every device and terminal in the near future, promising peak data rates up to 20Gb/s and average data rates greater than 100Mb/s. Energy efficiency (EE) becomes an important design criterion for the sustainable evolution of 5G networks [1] and tranceiver redesign offers great promise to improve EE but current literature is limited and further investigation is required [2]. The key enabler technologies to achieve these data rates are massive multiple-input multiple-output (MM) and the use of millimeter wave (mmWave) frequencies. Although promising, these approaches present several challenges that roadblock the design of energy efficient networks. On one hand is the a huge number of radio-frequency (RF) chains that have to be connected to the antennas. On the other is the large amounts of power required from the high resolution analog-to-digital converters (ADCs) under large-bandwidth operation with multi gigabit requirements.

Channel estimation of mmWave-based MM systems with few-bit ADCs poses major difficulties compared to conventional systems due to beamforming and the large number of antenna elements [3, 4]. Recently, techniques based on Approximate Message Passing (AMP) algorithm have been proposed for estimation of the MM with low-resolution ADCs [5, 6, 7]. In [6] a modification of the conventional Expectation-Maximization (EM) algorithm has been proposed which exploits the sparse modeling of the mmWave channel. However the technique degrades for the case of medium to high SNR regimes due to the non-linearities of the quantization. In [5] a generalized approximate message passing algorithm has been proposed that overcomes this divergence for large number of training length, however the same number of RF chains and antennas was assumed. In [7], an EM algorithm with the Iterative Hard Thresholding method has been proposed for the case of 1-bit ADCs.

In this work we consider the channel estimation problem for a mmWave-based massive MIMO with hybrid A/D beamforming, where the number of RF chains is lower than the number of antennas, while each chain employs a low-resolution ADC (1-3 bits). We introduce a *dithered beamforming* architecture where random control signals are added to the analog part of the beamformers and to the ADC prior the quantization of the input signal. Dithering has been successfully used for combat quantization effects [8], but never has been considered for massive MIMO beamforming. To recover the channel we consider the maximum a-posteriori estimator and we propose a *low-complexity* modification of the Expectation-Maximization (EM) algorithm which exploits the sparse representation of the mmWave channel into the beamspace (virtual) domain [9].

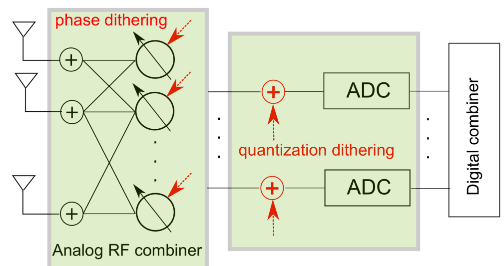

Fig. 1. Proposed dithered beamforming architecture. The random control signals are represented by red dashed arrows.

The authors acknowledge support from the Engineering and Physical Sciences Research Council (EPSRC) [EP/P000703/1].

2. PROBLEM FORMULATION

We consider the beamspace (virtual) representation of a MIMO channel with uniform linear arrays (ULA) [9, 10]. For simplicity we consider the narrowband case [11, 12], although the proposed architecture can be extended to a frequency-selective scenario [5]. For M_t transmit antennas and M_r receive antennas, the $M_r \times M_t$ channel matrix $\mathbf{H} \in \mathbb{C}^{M_r \times M_t}$ can be expressed as

$$\mathbf{H} = \mathbf{U}_r \mathbf{Z} \mathbf{U}_t^H = \sum_{i=1}^{M_t} \sum_{i=1}^{M_r} [\mathbf{Z}]_{ik} \mathbf{a}_r(\phi_i) \mathbf{a}_t^H(\theta_k) \tag{1}$$

where $\mathbf{a}_t(\theta_k) = \frac{1}{\sqrt{M_t}}[1, e^{-j\theta_k}, \dots, e^{-j(i-1)\theta_k}]^T$ is the steering vector of the transmitter with $\theta_k = k/M_t$ the normalized uniformly spaced spatial angles. The matrices $\mathbf{U}_r \in \mathbb{C}^{M_r \times M_r}$, $\mathbf{U}_t \in \mathbb{C}^{M_t \times M_t}$ are DFT matrices while $\mathbf{Z} \in \mathbb{C}^{M_r \times M_t}$ is a sparse matrix with a few non-zero elements. Specifically, each element of the sparse matrix $[\mathbf{Z}]_{ij}$ is assumed to follow the Bernoulli-Gaussian distribution, i.e.,

$$p([\mathbf{Z}]_{ij}) = (1-\eta)\delta([\mathbf{Z}]_{ij}) + \frac{\eta}{\sqrt{2\pi}\sigma_h} e^{-\frac{|[\mathbf{Z}]_{ij}|^2}{2\sigma_h^2}}$$

where $\delta(\cdot)$ is the Dirac delta function and $\eta = \frac{L}{M_t M_r}$ denotes the sparsity of virtual channel.

In the hybrid analog/digital (A/D) beamforming architecture, the number of transmitter RF chains M_t^{rf} is usually smaller than the number of the transmitting antennas M_t and similarly for the receiver $M_r^{rf} < M_r$. At each training instance t, the transmitter generates the vector $\mathbf{s}(t) \in \mathbb{C}^{M_t^{rf} \times 1}$, which is the input to the analog RF precoder, $\mathbf{F}_{RF}(t) \in \mathbb{C}^{M_t^{rf} \times M_t}$. This signal is transmitted through the sparsely modeled channel $\hat{\mathbf{H}}$ and the received vector is processed by the analog RF combiner $\mathbf{W}_{RF}(t) \in \mathbb{C}^{M_r \times M_r^{rf}}$. The output of the combiner can be written as:

$$\mathbf{y}_c(t) = \big(\underbrace{\mathbf{s}^T(t)\mathbf{F}_{RF}^T(t)\mathbf{U}_t \otimes \mathbf{W}_{RF}^H(t)\mathbf{U}_r}_{\mathbf{\Psi}_c(t) \in \mathbb{C}^{M_r \times M_r M_t}}\big)\mathbf{z}_c + \mathbf{W}_{RF}^H(t)\mathbf{n}_c(t) \tag{2}$$

where $\mathbf{z}_c = vec(\mathbf{Z})$, $\mathbf{n}_c \in \mathbb{C}^{M_r^{rf} \times 1}$ and $\mathbf{n}_c \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I}_{M_r})$, is the additive white complex Gaussian noise (AWGN) vector i.e., is generated based on the complex normal vector distribution. By concatenating all the T training sequences the into the real-valued equivalent form we have:

$$\bar{\mathbf{y}} = \begin{bmatrix} \mathrm{Re}(\bar{\mathbf{y}}_c) \\ \mathrm{Im}(\bar{\mathbf{y}}_c) \end{bmatrix} = \bar{\mathbf{\Psi}} \begin{bmatrix} \mathrm{Re}(\mathbf{z}_c) \\ \mathrm{Im}(\mathbf{z}_c) \end{bmatrix} + \begin{bmatrix} \mathrm{Re}(\bar{\mathbf{n}}_c) \\ \mathrm{Im}(\bar{\mathbf{n}}_c) \end{bmatrix} \tag{3}$$

where $\bar{\mathbf{\Psi}} = \begin{bmatrix} \mathrm{Re}(\bar{\mathbf{\Psi}}_c) & -\mathrm{Im}(\bar{\mathbf{\Psi}}_c) \\ \mathrm{Im}(\bar{\mathbf{\Psi}}_c) & \mathrm{Re}(\bar{\mathbf{\Psi}}_c) \end{bmatrix}^T \in \mathbb{R}^{2TM_r \times 2M_r M_t}$ and $\bar{\mathbf{y}}_\mathbf{c} \in \mathbb{C}^{TM_r \times 1}$, $\bar{\mathbf{n}}_\mathbf{c} \in \mathbb{C}^{TM_r^{rf} \times 1}$, $\bar{\mathbf{\Psi}}_\mathbf{c} \in \mathbb{C}^{TM_r \times M_r M_t}$ are the concatenated complex quantities for the received signal, the AWGN and the system matrix, respectively.

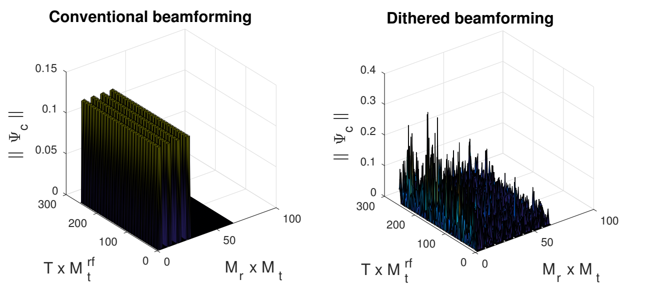

Fig. 2. Removal of the large scale patterns of the system matrix $\|\mathbf{\Psi}_c\|$ for $M_t = M_r = 8$, $M_t^{rf} = M_r^{rf} = 4$, $T = 64$.

We consider that the mmWave-based massive MIMO receiver employs low-resolution quantization at the ADCs. Let us denote the K-level quantization of $\bar{\mathbf{y}} \in \mathbb{R}^{2TM_r \times 1}$ as

$$\bar{\mathbf{q}} = \mathcal{Q}(\bar{\mathbf{y}}) \tag{4}$$

where $\bar{\mathbf{q}} = [q_1 \dots q_{2TM_r}]^T \in \mathbb{R}^{2TM_r \times 1}$. Each output element takes one of the K distinct values i.e., $q_i^1, \dots, q_i^K$. with $q_i^k = -(M+1) + k\Delta$ depending on the quantizer lower and upper thresholds $[l_i^k, u_i^k]$. The lower and upper quantizer boundary values are set to $q_{min} = -\kappa\sqrt{\mathcal{E}\{y_i^2\}}$ and $q_{max} = \kappa\sqrt{\mathcal{E}\{y_i^2\}}$, $\forall i$ and for $\kappa \in [1, 5]$, respectively. The quantizer's step-size is given by $\Delta = \frac{q_{max} - q_{min}}{M}$, while the average power $\mathcal{E}\{y_i^2\}$ can be obtained via an automatic gain control circuit.

Before proceeding, let us describe in more detail two main issues that render the channel estimation problem more challenging in the case of channel estimation at the receiver of a massive MIMO with hybrid A/D beamforming and low-resolution quantization. The first issue comes from the channel subspace sampling limitation [13] which prevents the direct estimation of the CSI due to the beamforming matrices. In the conventional case, where the beamforming matrices are composed by DFT columns, the resulting measurement matrix $\mathbf{\Psi}_c$ has a block structure with areas of similar values, as shown in Fig. 2 (left). This implies that $\mathrm{rank}(\mathbf{\Psi}_c) = \mathrm{rank}(\mathbf{W}_{RF}^H(t)\mathbf{U}_r) \leq M_t M_r$ and infinite condition number. Moreover, taking into account the quantization of the received signal, the overall system, given by (4), is a non-linear one due to the staircase ADCs, especially for the low-resolution cases (1-3 bits).

3. DITHERED BEAMFORMING

To overcome the aforementioned problems for channel estimation, we introduce a novel architecture, termed as *dithered beamforming*. Dithering is a commonly used technique where an external signal is injected to the input to combat the non-linear quantization effects, improve the robustness and the asymptotic stability of the system [14, 15]. In our design, two external signals are injected at the MIMO receiver, one in the spatial angles and another in the amplitude, as shown in Fig. 1. Therefore, the use of dithering is two-fold: first we improve the properties of the measurement matrix by introducing randomness into the signal capturing process. Afterwards, the outputs of the RF combiner are perturbed by adding random analog memory-less signals to overcome the stair-case effects of low-resolution ADCs.

Let us consider that the steering vectors of precoding/combiner matrices have quantized physical angles which are generated as random variables following the uniform distribution. The spatial angle of the k-th antenna element the transmitter will be expressed as:

$$\theta_k = 2\pi \sin(Q_{K_\omega}(\omega_k)) \tag{5}$$

where $Q_{K_\omega}(\cdot)$ denotes a K_ω-level quantizer and $\omega_k \sim \mathcal{U}(0, 2\pi)$. Then, for each training instance t at the k-th antenna and the i-th RF chain we use the precoding element

$$[\mathbf{F}_{RF}(t)]_{ki} = \frac{1}{\sqrt{M_t}} e^{j(i-1)\sin(\theta_k^t(t))} \tag{6}$$

and accordingly for the combiner at the receiver. Thus, the resulting measurement matrix $\bar{\mathbf{\Psi}}$ will be stochastic and the block structure is removed [16], as shown in Fig. 2 (right). This is similar to the compressive sensing approach where a random measurement matrix is used to recover the sparse signal based on small number of measurements [17].

For the amplitude dithering, several approaches have been proposed for the signle-input-single-output case (SISO), e.g., random or deterministic, non-subtractive or subtractive [15]. In this work we consider a non-subtractive random dithering for a MIMO system, $\bar{\mathbf{d}} \in \mathbb{R}^{2TM_r^{rf} \times 1} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_d)$, hence, the overall system is described as:

$$\bar{\mathbf{r}} = \mathcal{Q}\big(\bar{\mathbf{\Psi}}\mathbf{z} + \bar{\mathbf{n}} + \bar{\mathbf{d}}\big) \in \mathbb{R}^{2TM_r \times 1} \tag{7}$$

where the overall noise can be modelled as $\bar{\mathbf{n}} + \bar{\mathbf{d}} \sim \mathcal{N}\big(\mathbf{0}, \mathbf{\Sigma}\big)$, where $\mathbf{\Sigma} = \sigma_n^2 \mathbf{W}_{rf}^H \mathbf{W}_{rf} + \mathbf{\Sigma}_d$.

Note that the proposed formulation introduces dithering for a MIMO system where the covariance matrix $\mathbf{\Sigma}$ determines its performance. Up to the authors knowledge this is the first time where correlated MIMO dithering is considered and extended investigation is left for the subsequent works.

It can be proved that when the number of quantized angles is large (e.g., $K_\omega > 8$), the restricted isometry property [17] for the beamforming matrices is satisfied, thus the approximation $\mathbf{W}_{rf}^H \mathbf{W}_{rf} \approx \mathbf{I}_N$ holds for large enough K_ω. To this end, the MIMO dithering system can be decomposed into independent SISO dithering systems. For the SISO case with random dithering, the dither variance has a crucial impact on the performance of dithered quantization and its optimal value depends on the bit resolution and on the dynamic range Δ of the quantizer [15].

4. LOW-COMPLEXITY CHANNEL ESTIMATOR

In this section we develop a low-complexity channel estimator for the dithered beamforming architecture. We consider the case where $K_\omega = 8$, hence random dithering can be described using a common parameter, i.e., the variance σ_d^2. To recover the CIR based on the non-linear input/output relation of (7), we provide a modified EM algorithm which approximates the maximum a-posteriori (MAP) estimator. The EM algorithm solves the following problem:

$$\mathcal{E}_{\bar{\mathbf{y}}|\bar{\mathbf{r}},\mathbf{z}} \left\{ \frac{\partial}{\partial \mathbf{z}} \ln p(\bar{\mathbf{r}}, \bar{\mathbf{y}}|\mathbf{z}^l) \right\} = 0 \tag{8}$$

where the conditional probability density function (PDF) involving $\bar{\mathbf{r}}$ and $\bar{\mathbf{y}}$ random variables is given by $p(\bar{\mathbf{r}}, \bar{\mathbf{y}}|\mathbf{z}) = \mathbb{I}_{D(\bar{\mathbf{r}})}(\bar{\mathbf{y}}) \frac{1}{(2\pi\sigma^2)^{N/2}} e^{\frac{-\|\bar{\mathbf{y}} - \bar{\mathbf{\Psi}}\mathbf{z}\|_2^2}{2\sigma^2}}$ [18]. EM algorithm is defined by the following two steps for the $(l+1)$-th iteration (details are omitted due to space limitations):

- **E-step:** Compute $\mathbf{b}^l = [b_1^l, \dots, b_{2TM_r}^l]$ with

$$b_i^l = -\frac{\sigma}{\sqrt{2\pi}} \frac{e^{-\frac{(l_i - [\bar{\mathbf{\Psi}}\mathbf{z}^l]_i)^2}{2\sigma^2}} - e^{-\frac{(u_i - [\bar{\mathbf{\Psi}}\mathbf{z}^l]_i)^2}{2\sigma^2}}}{\mathrm{erf}\big(\frac{-l_i + [\bar{\mathbf{\Psi}}\mathbf{z}^l]_i}{\sqrt{2}\sigma}\big) - \mathrm{erf}\big(\frac{-u_i + [\bar{\mathbf{\Psi}}\mathbf{z}^l]_i}{\sqrt{2}\sigma}\big)} \tag{9}$$

  where l_i, u_i are the lower/upper bounds of the quantizer for $[\bar{\mathbf{\Psi}}\mathbf{z}^l]_i$ respectively; $\mathrm{erf}(\cdot)$ is the error function.

- **M-step:** Estimate the sparse channel $\mathbf{z}^{l+1} \in \mathbb{R}^{2M_rM_t \times 1}$ via solution of the linear system of equations:

$$\mathbf{A}\mathbf{z}^{l+1} = \boldsymbol{\beta}_l \tag{10}$$

  with $\boldsymbol{\beta}_l \triangleq \bar{\mathbf{\Psi}}^T \bar{\mathbf{\Psi}} \mathbf{z}^l + \mathbf{b}^l$ and $\mathbf{A} \triangleq \bar{\mathbf{\Psi}}^T \bar{\mathbf{\Psi}} + \mathbf{C}_h^{-1}$.

The performance of EM algorithm is determined by the solver of (10). Given that prior PDF of the CSI, i.e., $p([\mathbf{Z}]_{ij})$, is known, several sparse solvers can be employed for the estimation of $\mathbf{z}$, e.g., AMP [19], CoSaMP [20], SGP [21], offering trade-offs between complexity, performance and prior knowledge. Since the matrix dimensions are expected to be very large in the massive MIMO case, matrix inversion is prohibitively complex.

On this premise, we employ an *approximate subspace gradient pursuit* technique which has low computational complexity with comparable performance, given that the channel sparsity η is known. At the $(l+1)$-th iteration of the EM algorithm, the support set Ω of the sparse channel vector is updated based on the approximated gradient vector $\mathbf{g}^{l+1}$. Specifically, Ω is the union of the previously selected one and the index which represents the element of the gradient with the highest energy. Note that by $\mathbf{A}_{|\Omega}$ we denote the submatrix which is obtained keeping the columns of $\mathbf{A}$ with indices in Ω. Afterwards, the sparse vector is approximated based on a line search update, where $\mathbf{t}$ is an auxiliary vector representing the pruned version of $\mathbf{z}$ (for further details see [21]).

The proposed low-complexity EM technique is presented in Algorithm 1 in more detail. The complexity order of the proposed algorithm is determined by the matrix-vector product of line 6, which is $\mathcal{O}(M_rM_tL)$. The number of non-zero

Algorithm 1 Proposed low-complexity EM algorithm

Input: $\beta, \bar{\mathbf{\Psi}}, \mathbf{C}_h, \sigma, u_i, l_i$
Output: $\mathbf{z}$

Initialization: $\mathbf{t} = \mathbf{0}$, $\mathbf{z} = \mathbf{0}$, $\mathbf{A} = \bar{\mathbf{\Psi}}^T \bar{\mathbf{\Psi}} + \mathbf{C}_h^{-1}$
for $l = 1, \ldots$ **do**
  // Expectation step
  Estimate $\mathbf{b}^l$ via (9)
  // Update the approximate gradient
  $\mathbf{g}^{l+1} = \boldsymbol{\beta}^l - \mathbf{A}_{|\Omega} \mathbf{t}^l_{|\Omega}$
  // Update the support set
  $\Omega = \arg\max(|\mathbf{z}(n-1)|, L-1) \cup \arg\max(|\mathbf{g}^l|, 1)$
  // Approximate the channel vector
  $\alpha = \frac{(\mathbf{g}^l_{|\Omega})^T \mathbf{g}^l_{|\Omega}}{(\mathbf{g}^l_{|\Omega})^T \mathbf{A}_{|\Omega} \mathbf{g}^l_{|\Omega}}$
  $\mathbf{z}^{l+1}_{|\Omega} = \mathbf{t}^l_{|\Omega} + \alpha \mathbf{g}^l_{|\Omega}$
  // Prune the non-zero values
  $\mathbf{t}^{l+1}_{|\Omega} = \mathbf{z}^{l+1}_{|\Omega}$
  $\mathbf{t}^{l+1}_{|\Omega^c} = \mathbf{0}$
end for

coefficients in the sparse vector $\mathbf{t}$ is L which is much lower than the number of the antennas $M_t M_r$. Note that the complexity order for the matrix inversion would require two orders of magnitude higher complexity, i.e., $\mathcal{O}((M_r M_t)^3)$.

5. SIMULATION RESULTS AND CONCLUSION

In this section, we evaluate the performance of the proposed algorithm through simulation results. We assume that the transmitter and the receiver have $M_t = M_r = 32$ antennas and $M_t^{rf} = M_r^{rf} = 8$ RF chains. The training sequence $\mathbf{s}(t) \times \mathbb{C}^{M_t \times 1}, t = 1, \ldots, T$ is composed of 4-QAM symbols with $\mathcal{E}\{\|\mathbf{s}(t)\|_2^2\} = 1$. We assume ULA antenna arrays with $\lambda/2$ spacing. The mmWave channel is considered to be quasi-static, i.e., it remains static during the transmission of the $T = 512$ training vectors. The number of channel paths was set to $L = 40$, i.e., the sparsity level is $\eta \sim 1\%$, and its variance to $\sigma_z^2 = L$. The dithering variance σ_d^2 has been set to the optimal value for each bit resolution, and is obtained by exhaustive search.

In Fig. 3 we evaluate the convergence behaviour of the proposed algorithm (Algorithm 1) depicted in for two SNR cases, i.e., 0dB and 15dB. To evaluate the steady-state performance we have also plotted the oracle-based EM algorithm, where (10) is solved by using least-squares with known support set Ω, i.e., $\mathbf{z}^{oracle}_{|\Omega} = \mathbf{A}^{-1}_{|\Omega} \boldsymbol{\beta}$. We see that the proposed algorithm is able to achieve very rapid convergence rates for both SNR values, despite only using an approximate solution for (10). For the 1-bit case, the MSE performance of the proposed technique is almost optimal, while for 2 and 3 bits cases is 1dB higher than the optimal.

In Fig. 4 we compare the MSE with respect to the SNR for dithered and non-dithered cases. Apart from the proposed one, we also have included an AMP-based solution of (10) [19]. Channel estimation without dithering for SNRs over 10dB exhibits very large errors due to non-linear phenomena of the quantization. However, the proposed dithered beamforming technique is able to mitigate these effects. Note also that the proposed algorithm achieves superior performance to the AMP technique; AMP is more appropriate for the cases where the $|\Omega|$ is unknown and has to be also recovered.

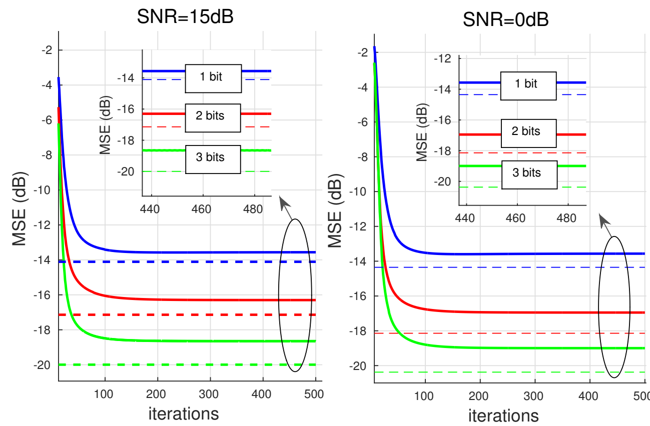

Fig. 3. Convergence curves for the proposed Algorithm 2.

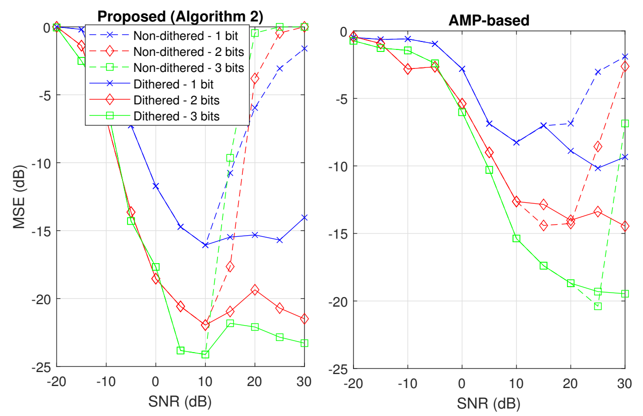

Fig. 4. MSE for dithered and non-dithered quantization cases.

In conclusion, this work has considered channel estimation at the receiver of a massive MIMO system with hybrid A/D beamforming and low-resolution quantization. A dithered beamforming architecture has been introduced where random control signals are added to the analog part of the combiner and to the ADC prior to the quantization of the input signal. A low-complexity EM algorithm has been proposed to exploit the sparse representation of the mmWave channel. From the results of our study we conclude that the proposed dithered beamforming architecture is able to mitigate the effects of ill-condition of the system matrix as well as the non-linear effects of quantization that occur during channel estimation. Also, the proposed algorithm can achieve almost optimal MSE performance with low complexity requirements.

6. REFERENCES

[1] J. Thompson, X. Ge, H. C. Wu, R. Irmer, H. Jiang, G. Fettweis, and S. Alamouti, "5G wireless communication systems: prospects and challenges [Guest Editorial]," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 62–64, Feb. 2014.

[2] K. N. R. S. V. Prasad, E. Hossain, and V. K. Bhargava, "Energy efficiency in massive MIMO-based 5G networks: Opportunities and challenges," *IEEE Wireless Communications*, vol. 24, no. 3, pp. 86–94, Jun. 2017.

[3] R. W. Heath, N. Gonzlez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[4] M. Xiao, S. Mumtaz, Y. Huang, L. Dai, Y. Li, M. Matthaiou, G. K. Karagiannidis, E. Björnson, K. Yang, C. L. I, and A. Ghosh, "Millimeter wave communications for future mobile networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 1909–1935, Sep. 2017.

[5] J. Mo, P. Schniter, and R. W. Heath, "Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs," *IEEE Transactions on Signal Processing*, vol. 66, no. 5, pp. 1141–1154, Mar. 2018.

[6] J. Mo, P. Schniter, N. G. Prelcic, and R. W. Heath, "Channel estimation in millimeter wave MIMO systems with one-bit quantization," in *2014 48th Asilomar Conference on Signals, Systems and Computers*, Nov. 2014, pp. 957–961.

[7] C. Stckle, J. Munir, A. Mezghani, and J. A. Nossek, "Channel estimation in massive MIMO systems using 1-bit quantization," in *2016 IEEE 17th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Jul. 2016, pp. 1–6.

[8] G. Zeitler, G. Kramer, and A. C. Singer, "Bayesian parameter estimation using single-bit dithered quantization," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 2713–2726, Jun. 2012.

[9] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 7, pp. 3814–3827, Jul. 2013.

[10] L. Dai, X. Gao, S. Han, I. Chih-Lin, and X. Wang, "Beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array," in *2016 IEEE/CIC International Conference on Communications in China (ICCC)*, July 2016, pp. 1–6.

[11] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[12] R. Mndez-Rial, C. Rusu, N. Gonzlez-Prelcic, A. Alkhateeb, and R. W. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?," *IEEE Access*, vol. 4, pp. 247–267, 2016.

[13] T. Kim and D. J. Love, "Virtual AoA and AoD estimation for sparse millimeter wave MIMO channels," in *2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2015, pp. 146–150.

[14] F.A. Malekzadeh, R. Mahmoudi, and A. van Roermund, *Analog Dithering Techniques for Wireless Transmitters*, Analog Circuits and Signal Processing. Springer New York, 2012.

[15] H. C. Papadopoulos, G. W. Wornell, and A. V. Oppenheim, "Sequential signal encoding from noisy measurements using quantizers with dynamic bias control," *IEEE Transactions on Information Theory*, vol. 47, no. 3, pp. 978–1002, Mar. 2001.

[16] Gene H. Golub and Charles F. Van Loan, *Matrix Computations (3rd Ed.)*, Johns Hopkins University Press, Baltimore, MD, USA, 1996.

[17] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theor.*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[18] O. Dabeer and E. Masry, "Multivariate signal parameter estimation under dependent noise from 1-bit dithered quantized data," *IEEE Transactions on Information Theory*, vol. 54, no. 4, pp. 1637–1654, Apr. 2008.

[19] David L. Donoho, Arian Maleki, and Andrea Montanari, "Message passing algorithms for compressed sensing," *CoRR*, vol. abs/0907.3574, 2009.

[20] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," Tech. Rep., California Institute of Technology, Pasadena, 2008.

[21] E. Vlachos, A. S. Lalos, and K. Berberidis, "Stochastic gradient pursuit for adaptive equalization of sparse multipath channels," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 2, no. 3, pp. 413–423, Sep. 2012.