Reduced Complexity Approximate Message Passing for Hybrid Architecture Based Millimeter Wave Massive MIMO Channel Estimation

Mariam Mussbah, Stefan Schwarz, and Markus Rupp
Institute of Telecommunications, Technische Universität Wien, Vienna, Austria
Christian Doppler Laboratory for Dependable Wireless Connectivity for the Society in Motion
Email: {mariam.mussbah, stefan.schwarz, markus.rupp}@tuwien.ac.at

***Abstract*—We develop a low complexity channel estimation algorithm for millimeter wave massive MIMO systems deploying a hybrid architecture. Motivated by the sparsity of the millimeter wave massive MIMO channel in the angular domain, many compressed sensing-based channel estimation algorithms have been developed. As the number of antennas grows, the size of the sensing matrix becomes very large, leading to high complexity and large storage requirements. In this paper, we develop a low complexity Bayesian approximate message passing based channel estimation scheme. The complexity and the storage requirements will be reduced by designing analog and digital precoders that result in a sparse sensing matrix, which achieves a recovery performance close to a full-size random Gaussian sensing matrix.**

***Index Terms*—Channel estimation, millimeter wave communication, hybrid precoding, massive MIMO, approximate message passing.**

I. Introduction

The deployment of large antenna arrays is critical to overcome the high path losses at the millimeter wave (mmWave) frequencies and to maintain sufficient link quality. To achieve the high beamforming gains, promised by massive multiple input multiple output (MIMO), channel state information (CSI) based directional beamforming is required. Due to the high cost and power consumption, the dedication of a separate radio frequency (RF) chain for each antenna is not suitable for massive MIMO. To reduce power consumption and costs, the separation of the precoding into an analog and digital precoder (hybrid precoding) [1], [2] and the use of low-resolution analog to digital converters [3] have attracted a lot of interest.

A common assumption for achieving the gains promised by mmWave massive MIMO systems is that CSI is available at the transmitter and/or receiver. Hybrid architecture based mmWave massive MIMO channel estimation algorithms do not directly observe the channel and operate at a low signal-to-noise ratio. Therefore, the conventional per-antenna channel estimation is not applicable. One way to reduce the training overhead is by exploiting the sparse scattering nature of the mmWave channel [4]. Several compressed sensing based channel estimation algorithms were proposed to strike a balance between training overhead and estimation accuracy [5]–[14]. However, most of these compressed sensing based solutions use large sensing matrices, resulting in high storage requirements and high complexity.

In this paper, we propose a low complexity Bayesian approximate message passing (BAMP) based channel estimation scheme. We reduce the size of the sensing matrix by first formulating a per user equipment (UE) antenna compressed sensing problem. The resulting sensing matrix is independent of the antenna index, resulting in a reduction of the required storage proportional to the squared number of UE antennas. Then, we will present a digital and analog precoders designing strategy that results in a sparse sensing matrix. The sparse design reduces the required storage and the complexity associated with the matrix-vector multiplication related to the sensing matrix, thus resulting in a reduction of the complexity of the BAMP algorithm. Since the full-size random Gaussian matrix achieves close to optimal performance [15], we will compare the recovery performance and complexity of the sparse matrix to the recovery performance and complexity of a full-size Gaussian sensing matrix.

The following notation is used throughout this paper: $\mathbf{A}$ is a matrix, $\mathbf{a}$ is a vector, a is a scalar and the i-th column of $\mathbf{A}$ is denoted by $\mathbf{a}_i$. The all-ones vector is denoted by $\mathbf{1}$. The transpose of a matrix is $\mathbf{A}^T$, the conjugate-transpose is $\mathbf{A}^H$ and the Frobenius norm is $\|\mathbf{A}\|_F$.

II. System Model

We consider a massive MIMO mmWave communication system in which the base station (BS) is equipped with N_{BS} antennas and the UE is equipped with N_{UE} antennas. The BS is equipped with M_{BS} RF chains and a fully connected phase shifters based hybrid precoding architecture. For the UE we assume a fully digital architecture. On the downlink, the BS first precodes the pilot symbol p, using the baseband precoder $\mathbf{f}_{\text{BB}} \in \mathbb{C}^{M_{\text{BS}} \times 1}$, followed by the RF precoder $\mathbf{F}_{\text{RF}} \in \mathbb{C}^{N_{\text{BS}} \times M_{\text{BS}}}$. Since the RF precoder is implemented using analog phase shifters and combiners, each element of $\mathbf{F}_{\text{RF}}$ is normalized to satisfy $|f_{\text{RF},i,j}| = \frac{1}{\sqrt{N_{\text{BS}}}}$. The total transmit power constraint is enforced by normalizing $\mathbf{f}_{\text{BB}}$ to satisfy $\|\mathbf{F}_{\text{RF}}\mathbf{f}_{\text{BB}}\|_{\text{F}}^2 = P_t$. Assuming block fading, the received signal at time instance k

can be modeled as

$$\mathbf{y}_k = \mathbf{W}_k \mathbf{H} \mathbf{F}_{\text{RF},k} \mathbf{f}_{\text{BB},k} p_k + \mathbf{W}_k \mathbf{n}_k, \tag{1}$$

where $\mathbf{H} \in \mathbb{C}^{N_{\text{UE}} \times N_{\text{BS}}}$ denotes the channel, $\mathbf{W}_k \in \mathbb{C}^{N_{\text{UE}} \times N_{\text{UE}}}$ is the digital combiner applied at the UE, and $\mathbf{n}_k \in \mathbb{C}^{N_{\text{UE}} \times 1}$ is the noise vector. The components of $\mathbf{n}_k$ are assumed to be independent and identically distributed (i.i.d.) circularly symmetric Gaussian $\mathcal{CN}(0, \sigma_n^2)$. Under the assumption that $\|p_k\|^2 = 1$, the received signal can be rewritten as

$$\tilde{\mathbf{y}}_k = \mathbf{y}_k p_k^* = \mathbf{W}_k \mathbf{H} \mathbf{F}_{\text{RF},k} \mathbf{f}_{\text{BB},k} + \mathbf{W}_k \tilde{\mathbf{n}_k}, \tag{2}$$

where $\tilde{\mathbf{n}_k} = \mathbf{n}_k p_k^*$ has the same statistics as $\mathbf{n}_k$.

For the channel, we adopt a geometric channel with N_c clusters to model the sparse scattering nature of the mmWave channels [4]. Furthermore, we assume that each scatterer contributes a single propagation path between the BS and the UE [16]. The channel $\mathbf{H}$ can be expressed by

$$\mathbf{H} = \sum_{c=1}^{N_c} \beta_c \mathbf{a}_r \left(\theta_c^{(r)}, \Phi_c^{(r)}\right) \mathbf{a}_t^T \left(\theta_c^{(t)}, \Phi_c^{(t)}\right), \tag{3}$$

where β_c is the complex gain of the c-th cluster, $\theta_c^{(r)}$ and $\Phi_c^{(r)}$ are the elevation and azimuth angles of arrival (AoA), and $\theta_c^{(t)}$ and $\Phi_c^{(t)}$ are the elevation and azimuth angles of departure (AoD). Finally, $\mathbf{a}_r \left(\theta_c^{(r)}, \Phi_c^{(r)}\right)$ and $\mathbf{a}_t \left(\theta_c^{(t)}, \Phi_c^{(t)}\right)$ are the array steering vectors at the receiver and transmitter side. For a uniform planar array (UPA), the array steering vector can be expressed as

$$\mathbf{a}(\theta) = \mathbf{a}_v(\theta) \otimes \mathbf{a}_h(\theta, \phi) \tag{4}$$

with

$$\begin{aligned} \mathbf{a}_v(\theta) &= \frac{1}{\sqrt{N_v}} \left[1, e^{-j\Psi_v}, ..., e^{-j\Psi_v(N_v-1)}\right] \\ \mathbf{a}_h(\theta, \phi) &= \frac{1}{\sqrt{N_h}} \left[1, e^{-j\Psi_h}, ..., e^{-j\Psi_h(N_h-1)}\right], \end{aligned} \tag{5}$$

where N_v and N_h are the number or rows and columns of the UPA, $\Psi_v = \frac{2\pi}{\lambda} d \cos(\theta)$, $\Psi_h = \frac{2\pi}{\lambda} d \sin(\theta)\cos(\phi)$, d denotes the inter-antenna spacing, and λ is the wavelength.

The angular domain representation of the channel in (3) is given by

$$\mathbf{H}_a = \mathbf{U}_r \mathbf{H} \mathbf{U}_t, \tag{6}$$

where $\mathbf{U}_r$ and $\mathbf{U}_t$ are the $N_{\text{UE}} \times N_{\text{UE}}$ and $N_{\text{BS}} \times N_{\text{BS}}$ discrete Fourier transform (DFT) matrices and their (m, n)-th element is given by

$$u_{t,m,n} = \frac{1}{\sqrt{N_{\text{BS}}}} e^{\frac{j2\pi}{N_{\text{BS}}}(m-1)(n-1)} \tag{7}$$

$$u_{r,m,n} = \frac{1}{\sqrt{N_{\text{UE}}}} e^{\frac{j2\pi}{N_{\text{UE}}}(m-1)(n-1)}, \tag{8}$$

Fig. 1 shows the angular domain representation of the channel matrix for $N_{\text{BS}} = 128$, $N_{\text{UE}} = 1$, $N_c = 4$. Further, we have assumed that the AoAs and AoDs are generated independently from $\mathcal{U}(0, \pi)$. To evaluate the angular domain representation, we show in Fig. 2 the total power of the strongest i components of $\mathbf{H}_a$. It reveals that about 10% of the components of $\mathbf{H}_a$ capture 90% of the power of the channel. Therefore, for large antenna arrays and at low signal to noise ratio (SNR), the angular domain representation can be modeled as approximately sparse.

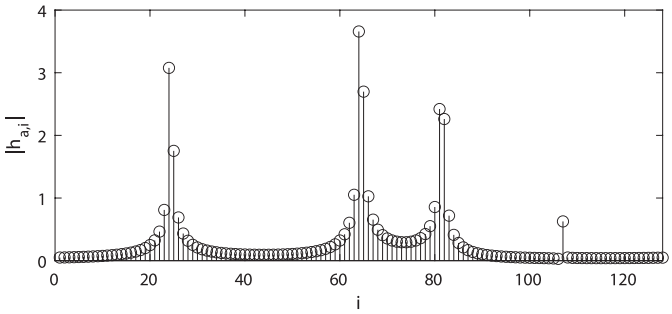

Fig. 1: Angular domain representation for $N_{\text{BS}} = 128$, $N_{\text{UE}} = 1$ and $N_c = 4$.

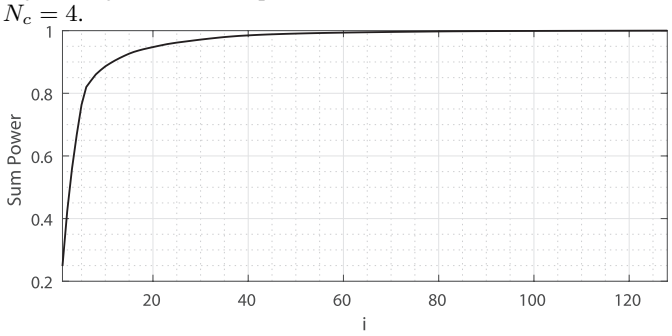

Fig. 2: The normalized sum power of the i strongest components of the angular domain representation.

By plugging in (6) in (2) and vectorizing the right hand side, the received signal can be rewritten as

$$\begin{aligned} \tilde{\mathbf{y}}_k &= (\mathbf{U}_t^H \mathbf{F}_{\text{RF},k} \mathbf{f}_{\text{BB},k})^T \otimes (\mathbf{W}_k \mathbf{U}_r^H) \text{vec}(\mathbf{H}_a) + \mathbf{W}_k \tilde{\mathbf{n}}_k \\ &= \mathbf{\Phi}_k \mathbf{h}_a + \tilde{\mathbf{n}}_k, \end{aligned} \tag{9}$$

where $\mathbf{h}_a = \text{vec}(\mathbf{H}_a)$ and $\mathbf{\Phi}_k = (\mathbf{U}_t^H \mathbf{F}_{\text{RF},k} \mathbf{f}_{\text{BB},k})^T \otimes (\mathbf{W}_k \mathbf{U}_r^H)$. Stacking the measurements obtained from sending K pilot vectors using different precoders and combiners, we get

$$\mathbf{y} = \mathbf{\Phi} \mathbf{h}_a + \tilde{\mathbf{n}}, \tag{10}$$

where $\mathbf{y} = \left[\mathbf{y}_1^T, \mathbf{y}_2^T, ..., \mathbf{y}_K^T\right]^T$ is the total received vector, $\mathbf{\Phi} = \left[\mathbf{\Phi}_1^T, \mathbf{\Phi}_2^T, ..., \mathbf{\Phi}_K^T\right]^T$ is the sensing matrix and $\tilde{\mathbf{n}} = \left[\mathbf{W}_1 \tilde{\mathbf{n}}_1^T, \mathbf{W}_2 \tilde{\mathbf{n}}_2^T, ..., \mathbf{W}_K \tilde{\mathbf{n}}_K^T\right]^T$ is the total noise vector. Estimating the channel simplifies now to estimating the approximately sparse vector $\mathbf{h}_a$ from the received signal vector $\mathbf{y}$.

Given the sparse representation of the channel, compressed sensing theory allows a stable recovery of the channel matrix with much fewer measurements than the total number of coefficients. In what follows, we will present the BAMP algorithm for recovering sparse signals from few measurements.

III. BAMP

Bayesian estimation is a method in which both, the unknown signal vector $\mathbf{x}$ and the additive noise vector are modeled as random, and it is assumed that prior information about the

statistic of $\mathbf{x}$ and the statistical relation between $\mathbf{x}$ and the measurements $\mathbf{y}$ is known. A well-known solution for the Bayesian estimator is the minimum mean square error (MMSE) estimator, which is given by $\hat{\mathbf{x}} = \mathbb{E}\{\mathbf{x}|\mathbf{y}\}$. BAMP [17], [18] is a reduced complexity sum-product algorithm that offers an approximate solution for the MMSE estimate for the following linear measurement model

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{n}, \tag{11}$$

where $\mathbf{\Phi}$ is the $K \times N$ sensing matrix and the components of the noise vector $\mathbf{n}$ are assumed to be i.i.d. zero-mean Gaussian with variance σ_n^2. The BAMP algorithm, including the initialization and stopping criterion, is stated in Algorithm 1. After incrementing the iteration counter in line 1, the effective noise variance estimate is calculated as the mean empirical power of the residual $\mathbf{z}^{(t-1)}$. In line 2, a decoupled measurement model is calculated. Next, the function $F(u_n^{(t-1)}; c^{(t-1)})$, which is parametrized by the effective noise variance $c^{(t-1)}$, acts on the decoupled measurements $u_n^{(t-1)}$ and calculates an estimate of $\mathbf{x}$. The algorithm stops if the relative change is smaller than ϵ or the maximum number of iterations t_{max} is reached.

Algorithm 1: BAMP

Input: $\mathbf{y}$, $\mathbf{\Phi}$, $t_{\max}$, ϵ
Initialization: $\hat{\mathbf{x}}^{(0)} = \mathbf{0}$, $\mathbf{z}^{(0)} = \mathbf{y}$, $t = 0$
do
1 $t = t + 1$
2 $c^{(t-1)} = \text{Var}\left(\mathbf{z}^{(t-1)}\right)$
3 $\mathbf{u}^{(t-1)} = \hat{\mathbf{x}}^{(t-1)} + \mathbf{\Phi}^T \mathbf{z}^{(t-1)}$
4 $\forall n \in \{1, 2, ..., N\}$: $\hat{x}_n^{(t)} = F(u_n^{(t-1)}; c^{(t-1)})$
5 $\mathbf{z}^{(t)} = \mathbf{y} - \mathbf{\Phi}\hat{\mathbf{x}}^{(t)} + \mathbf{z}^{(t-1)} \sum_{n=1}^{N} F'(u_n^{(t-1)}; c^{(t-1)})$
6 **while** $\left\|\hat{\mathbf{x}}^{(t)} - \hat{\mathbf{x}}^{(t-1)}\right\|_2 > \epsilon \left\|\hat{\mathbf{x}}^{(t)}\right\|_2$ and $t < t_{\max}$
Output: $\hat{\mathbf{x}} = \hat{\mathbf{x}}^{(t)}$

The estimator function is designed for the signal prior $f(x_n)$ and additive Gaussian noise and is given by

$$F(u_n; c) = \mathbb{E}_{x_n}\{x_n | u_n\} \tag{12}$$

$$F'(u_n; c) = \frac{\mathrm{d}}{\mathrm{d}u_n} F(u_n; c). \tag{13}$$

For massive MIMO systems, the dimension of the sensing matrix $\mathbf{\Phi}$ is expected to be very large. For example, if we assume a BS equipped with a 32×32 antenna array, a UE equipped with a 4×4 antenna array, and a sampling rate of $r = 0.5$, the sensing matrix will be of dimension $N = 32^2 \times 4^2 = 16384$ and $K = rN = 8192$. Storing the sensing matrix $\mathbf{\Phi}$ using 8-bytes for the complex-valued elements requires approximately 1 GB of memory. Further, as the number of antennas grows, the complexity of the matrix-vector multiplication in lines 3 and 5 in Algorithm 1 increases, resulting in overall higher complexity.

IV. Reduced Complexity BAMP

Since the UE is equipped with $M_{\text{UE}} = N_{\text{UE}}$ RF chains, we can perform a per antenna channel estimation. The received signal at antenna i can be written as

$$\tilde{y}_{k,i} = \mathbf{h}_i^T \mathbf{F}_{\text{RF},k} \mathbf{f}_{\text{BB},k} + \tilde{n}_{k,i}. \tag{14}$$

Following the same steps as in (9)

$$\begin{aligned} \tilde{y}_{k,i} &= (\mathbf{U}_t^H \mathbf{F}_{\text{RF},k} \mathbf{f}_{\text{BB},k})^T \mathbf{h}_{a,i} + \tilde{n}_{k,i}, \\ &= \boldsymbol{\phi}_k^T \mathbf{h}_{a,i} + \tilde{\mathbf{n}}_k, \end{aligned} \tag{15}$$

and by stacking the K measurements for the i-th received antenna we get the following measurement model

$$\tilde{\mathbf{y}}_i = \mathbf{\Phi}\mathbf{h}_{a,i} + \tilde{\mathbf{n}}_i, \tag{16}$$

where the sensing matrix $\mathbf{\Phi} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, ..., \boldsymbol{\phi}_K]^T$ is of dimension $K \times N_{\text{BS}}$ and is independent of the index of receive antenna. Performing a per receive antenna channel estimate has two advantages: first, it reduces the training overhead as it does not require applying different combiners at the receiver. And second, the size of the sensing matrix is reduced by a factor of N_{UE}^2. Compared to the example mentioned above, the memory required for saving the sensing matrix will be reduced from 1 GB to approximately 4 MB.

The matrix-vector multiplication for a sensing matrix of size $K \times N$ requires KN multiplications and $(K-1)N$ additions, resulting in approximately $2KN$ operations. In what follows, we will design the analog precoders $\mathbf{F}_{\text{RF},k}$ and the digital precoders $\mathbf{f}_{\text{BB},k}$ such that the resulting sensing matrix is sparse but still achieves recovery performance close to a full-size random Gaussian sensing matrix. The sparse sensing matrix has two advantages: first, it further reduces the required storage, and second, it reduces the complexity of the BAMP algorithm by reducing the complexity related to the matrix-vector multiplication.

In [19], the authors proposed a graph lifting based strategy for designing sparse sensing matrices. The sensing matrix has few randomly selected non-zero elements that take values from the set $\{-1, +1\}$. The authors start with a small $K_1 \times N_1$ prototype matrix of full-rank and with elements that randomly take the values $+1$ or -1. The sensing matrix is formed by repeating the prototype matrix according to the lifting factor L and randomly permuting the edges across the copies. The resulting sensing matrix is of size $LK_1 \times LN_1$ with only 1 in L matrix elements being non-zero. Further, they show that the recovery performance of the proposed sparse matrices is close to the recovery performance of the full-size Gaussian random sensing matrix. In Table I, we summarizes the commonly employed parameters.

From the measurement model (16), we know that for a sparse sensing matrix, each measurement corresponds to a linear combination of a small number of different elements of $\mathbf{h}_{a,i}$. To receive a linear combination of a small number of elements of $\mathbf{h}_{a,i}$, the analog and digital precoder have to be chosen such that the transmitted signal consists of a small number of beams that are steered into the direction of the

Symbol	Annotation
N_{BS}, N_{UE}	Number of BS and UE antennas
M_{BS}, M_{UE}	Number of BS and UE RF chains
N	Length of the unknown signal
K	Number of measurements
$K_1 \times N_1$	Size of the prototype matrix
L	Lifting factor

TABLE I: Frequently used parameters.

selected elements. The DFT-based codebook is well suited for steering beams [20]. Each column of the DFT matrix corresponds to a precoder that steers the beam into a different direction. Since the elements of the DFT matrix satisfy the constant modulus constraint, precoders from the DFT-based codebook can be realized by simple phase shifters. Therefore, the DFT-based codebook is well suited as a codebook for the analog precoders.

Given that the analog precoders for each RF chain are taken from a DFT-based codebook, the product $\mathbf{F}_{\text{BB},k}^T \mathbf{U}_t^H$ in (15) results in a sparse matrix which has in each row only one non-zero element. The position of the non-zero element in the i-th row is determined by the index of the i-th analog precoder, i.e., the index of the analog precoder which is used as the i-th column for $\mathbf{F}_{\text{BB},k}$. Further, left-multiplying the sparse matrix with the row vector $\mathbf{f}_{\text{BB},k}^T$ results in a sparse row vector for which the analog precoder determines the number of the non-zero elements and their position, and the digital precoder determines the weight of the non-zero elements. Consider, for example, a set-up with $N_{\text{BS}} = 8$, $M_{\text{BS}} = 3$ and a sensing matrix for which the k-th row is given by

$$\boldsymbol{\phi}_k = [0, 1, 0, 0, 1, -1, 0, 0], \tag{17}$$

the corresponding analog and digital precoders for the k-th transmission are given by

$$\mathbf{F}_{\text{RF},k} = [\mathbf{u}_{t,2}, \mathbf{u}_{t,5}, \mathbf{u}_{t,6}] \tag{18}$$

$$\mathbf{f}_{\text{BB},k} = [1, 1, -1]^T, \tag{19}$$

were $\mathbf{u}_{t,i}$ corresponds to the i-th column of the DFT matrix $\mathbf{U}_t$. In summary, the process of designing the analog and digital precoders consists of three steps: (1) a sparse matrix is designed based on [19]. The matrix is of dimension $K \times N_{\text{BS}}$ were each row corresponds to a transmission and has M_{BS} non-zeros elements. (2) The analog precoder is constructed from the DFT-based codebook. The indices of the columns of $\mathbf{U}_t$, which are used for constructing the analog precoder for the k-th transmission, are given by the indices of the non-zero elements in the k-th row of the sensing matrix. (3) The digital precoder for the k-th transmission is constructed based on the sign of the non-zero elements of the k-th row of the sensing matrix.

V. Numerical Results

In this section, we evaluate the performance of the proposed channel estimation scheme by running Monte-Carlo simulations. We set the number of clusters to $N_c = 4$, and for each cluster in (3), the corresponding gain is drawn independently from a zero-mean complex Gaussian distribution of variance $\sigma_{\beta_c}^2$. The channel power is normalized by enforcing $\sum_{c=1}^{N_c} \sigma_{\beta_c}^2 = 1$. For the normalized channel, the SNR is given by

$$\text{SNR} = \frac{P_t}{\sigma_n^2}. \tag{20}$$

The AoA and AoD are generated independently from $\mathcal{U}(0, \pi)$. We assume that the BS is equipped with a uniform linear array (ULA) of isotropic radiating antennas and half-wavelength spacing. Further, we assume that the UE is equipped with one antenna. We design the sparse sensing matrix with M_{BS} non-zero elements in each row according to [19] by using the following parameters: $N_1 = M_{\text{BS}}$, $L = N/M_{\text{BS}}$, and $K_1 = K/L$. As a prior distribution for the channel we use

$$f(x) = \gamma\delta(x) + (1-\gamma)\mathcal{N}(x; 0, \sigma_x^2), \tag{21}$$

where γ is the zero probability which models the sparsity in the angular domain, and the non-zero elements are modelled by a zero mean complex Gaussian distribution with variance σ_x^2. A derivation of the estimator function in (12) can be found in [21]. Further, we assumed the following parameters for the BAMP algorithm: $t_{\max} = 100$, $\epsilon = 10^{-4}$, and $\gamma = 0.8$.

We compare the recovery performance of the full-size matrix and the sparse sensing matrix in terms of normalized mean-square error (NMSE). The NMSE is defined as

$$\text{NMSE} = \mathbb{E}\left\{ \frac{\left\| \hat{\mathbf{H}} - \mathbf{H} \right\|_F^2}{\|\mathbf{H}\|_F^2} \right\}, \tag{22}$$

where $\hat{\mathbf{H}}$ denotes the estimated channel.

We evaluate the algorithm's computational complexity in terms of the required floating-point operations (FLOPs), where each operation is weighted according to its execution cost. The cost of each operation is given in terms of how much more expensive an operation is compared to the simplest operation, i.e., an addition. We summarize the weighting factors for the different operations in Table II [22].

	Weight
Multiplication	4
Division	10
Square root	30
Exponential function	50

TABLE II: Weighting factor for different operations.

In Table III, we summarize the total number of operations required per BAMP iteration for a full-size sensing matrix and a sparse sensing matrix. We have assumed that the total number of measurements is $K = rN_{\text{BS}}$. Weighting the operations according to Table II, we get a total of $O_f = 2rN_{\text{BS}}^2 + (102 + 5r)N_{\text{BS}} + 85$ FLOPs per iteration for the full-size matrix, and for the sparse matrix, we get $O_s = 10rN_{\text{BS}}M_{\text{BS}} + (102+5r)N_{\text{BS}} + 85$ FLOPs per iteration. We evaluate the complexity reduction by comparing the total number of FLOPs for the sparse matrix to the total number of FLOPs for the full-size matrix, i.e.,

$$\eta = \frac{t_s O_s}{t_f O_f}, \tag{23}$$

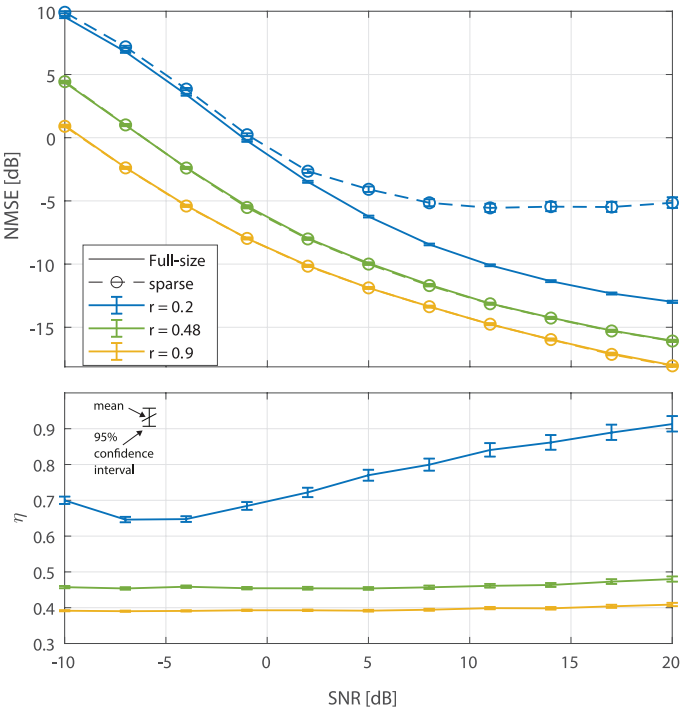

Fig. 3: Comparison of the performance and complexity of the BAMP algorithm for full-size and sparse sensing matrices.

where t_s and t_f denote the total number of iterations for the sparse sensing matrix and the full-size sensing matrix.

In Fig. 3, we compare the performance and complexity of the BAMP algorithm for a full-size and the sparse sensing matrix for $N_{\text{BS}} = 512$, $M_{\text{BS}} = 32$, and different sampling rates. We can see that the sparse matrix reduces the complexity by more than 50% for sufficiently high sampling rates while still achieving almost the same recovery performance as the full-size matrix. For low sampling rates, on the other hand, the sparse sensing matrix results in a higher error floor, and it requires more iteration resulting in a smaller overall complexity reduction.

In Fig. 4, we compare the performance for different degrees of sparsity for $N_{\text{BS}} = 512$ and $r = 0.2$. We can see that for low rates and a high degree of sparsity, the BAMP algorithm fails to recover the channel. As we decrease the sparsity of the sensing matrix, the difference in performance compared to the full-size sensing matrix decreases. For a lifting factor of $L = 8$, which is equivalent to a sparsity degree of 87.5%, the performance is already very close to that of the full-size sensing matrix. Thus, a high degree of sparsity comes at the cost of either a decreased performance or an increased overhead.

In our simulations, we have assumed that the angular domain representation is sparse, and we modeled it using a Bernoulli-Gaussian distribution. So far, we have drawn the AoAs and AoDs from a continuous uniform distribution, resulting in an approximate sparse angular presentation. To evaluate the effect of the sparse model assumption, we compare in Fig. 5 the recovery performance of the BAMP algorithm for an approximate sparse channel and a sparse channel. For the

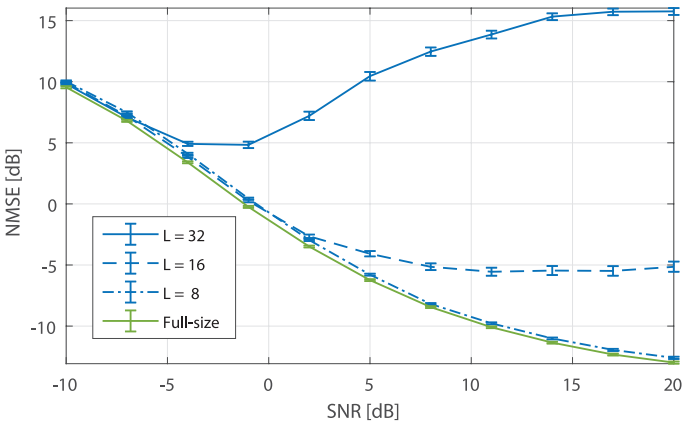

Fig. 4: Comparison of the performance of the BAMP algorithm for different lifting factors.

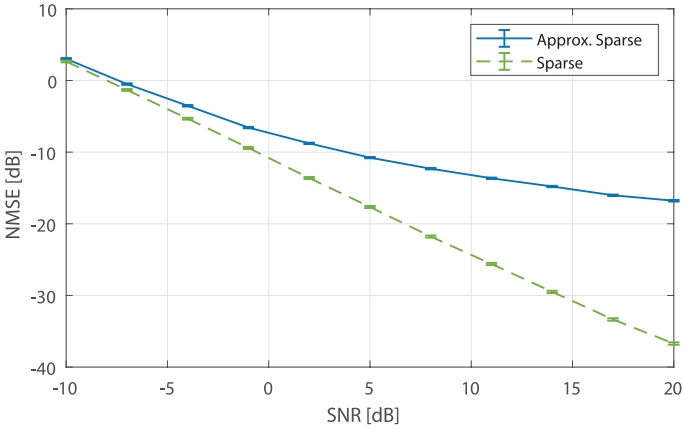

Fig. 5: Comparison of the performance of the BAMP algorithm for a sparse channel and an approximately sparse channel.

sparse channel, the angles are drawn from a discrete distribution. The discrete distribution is chosen such that the angles align with the beams from the codebook; therefore, resulting in a sparse angular domain representation. Fig. 5 shows that for low SNR, the sparse approximation models our channel quite well. At high SNR the NMSE starts saturating, whereas for the discrete angles the NMSE continues decreasing linearly with increasing SNR. The sparsity assumption does not hold for high SNR but the BAMP algorithm still tries to recover a sparse vector; therefore, only the strongest components are detected. Since the algorithm does not detect the weaker components, its performance saturates, and a further increase in SNR does not yield any significant improvement.

In Fig. 6, we compare the achievable rate for $N_{\text{BS}} = 512$, $M_{\text{BS}} = 32$, $r = 0.5$, and $U = 8$ users. We calculate the achievable rate as

$$R = \frac{1}{U} \sum_{u=1}^{U} \log_2 \left(1 + \frac{\left| \mathbf{h}_u^H \mathbf{F}_{\text{RF}} \mathbf{f}_{\text{BB},u} \right|^2}{\sum\limits_{i \neq u} \left| \mathbf{h}_u^H \mathbf{F}_{\text{RF}} \mathbf{f}_{\text{BB},i} \right|^2 + \sigma_n^2} \right). \quad (24)$$

Further, we assumed a phased zero forcing hybrid precoder [23] which is calculated based on the estimated channel. We observe that neglecting the weak components for low SNR

	Addition	Multiplication	Division	Square-root	Exponential
Full-size	$2rN_{\mathrm{BS}}^2 + (6+r)N_{\mathrm{BS}} - 1$	$3rN_{\mathrm{BS}}^2 + (9+r)N_{\mathrm{BS}} + 4$	$N_{\mathrm{BS}} + 4$	1	N_{BS}
Sparse	$2rN_{\mathrm{BS}}M_{\mathrm{BS}} + (6+r)N_{\mathrm{BS}} - 1$	$2rN_{\mathrm{BS}}M_{\mathrm{BS}} + (9+r)N_{\mathrm{BS}} + 4$	$N_{\mathrm{BS}} + 4$	1	N_{BS}

TABLE III: Required number of operations for a full-size matrix and a sparse matrix.

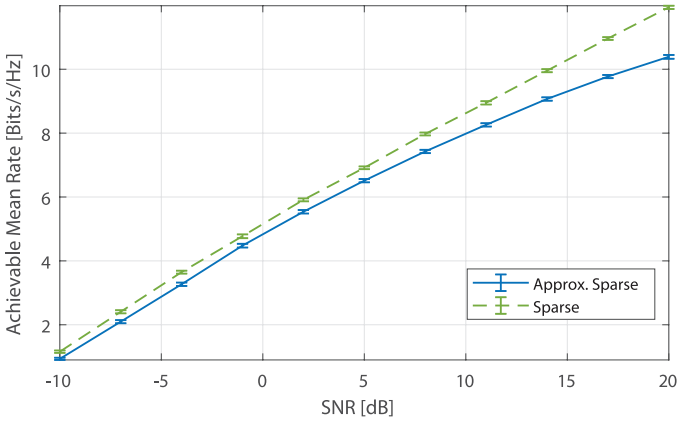

Fig. 6: Comparison of the achievable rate of the BAMP algorithm for a sparse signal and an approximately sparse signal.

does not have a big impact on the achievable rate. At high SNR, i.e., when the weak components exceed the noise level, we observe an increased degradation in the performance. Thus, the sparse assumption is a good approximation for low SNR and large antenna arrays and yields good results.

VI. Conclusion

In this paper, we have presented a low complexity BAMP algorithm for hybrid architecture based massive MIMO channel estimation. We reduce the overall complexity of the algorithm by applying a designing strategy for the precoders that results in a sparse sensing matrix. We have shown that with the sparse design, the estimation performance is close to the estimation performance of the full-size matrix while substantially reducing the storage requirements and the computational complexity.

Acknowledgements

The financial support by the Austrian Federal Ministry for Digital and Economic Affairs, the National Foundation for Research, Technology and development and the Christian Doppler Research Association is gratefully acknowledged

References

[1] S. Han, C. I, Z. Xu, and C. Rowell, "Large-scale Antenna Aystems with Hybrid Analog and Digital Beamforming for Millimeter Wave 5G," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 186–194, 2015.

[2] J. A. Zhang, X. Huang, V. Dyadyuk, and Y. J. Guo, "Massive Hybrid Antenna Array for Millimeter-Wave Cellular Communications," *IEEE Wireless Communications*, vol. 22, no. 1, pp. 79–87, 2015.

[3] L. N. Ribeiro, S. Schwarz, M. Rupp, and A. L. F. de Almeida, "Energy Efficiency of mmWave Massive MIMO Precoding With Low-Resolution DACs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 2, pp. 298–312, 2018.

[4] 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 Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1164–1179, 2014.

[5] C. Qi, Y. Huang, S. Jin, and L. Wu, "Sparse Channel Estimation Based on Compressed Sensing for Massive MIMO Systems," in *2015 IEEE International Conference on Communications (ICC)*, 2015, pp. 4558–4563.

[6] M. Chai and W. Zhou, "Overhead-efficient Channel Estimation and Beamforming for Hybrid Architecture-based mmWave Systems," in *2021 IEEE Wireless Communications and Networking Conference (WCNC)*, 2021, pp. 1–6.

[7] C. Huang, L. Liu, C. Yuen, and S. Sun, "Iterative Channel Estimation Using LSE and Sparse Message Passing for MmWave MIMO Systems," *IEEE Transactions on Signal Processing*, vol. 67, no. 1, pp. 245–259, 2019.

[8] J. Sung and B. L. Evans, "Hybrid beamformer codebook design and ordering for compressive mmwave channel estimation," in *2020 International Conference on Computing, Networking and Communications (ICNC)*, 2020, pp. 914–919.

[9] F. Bellili, F. Sohrabi, and W. Yu, "Generalized Approximate Message Passing for Massive MIMO mmWave Channel Estimation With Laplacian Prior," *IEEE Transactions on Communications*, vol. 67, no. 5, pp. 3205–3219, 2019.

[10] J. Rodríguez-Fernández, N. González-Prelcic, K. Venugopal, and R. W. Heath, "Frequency-Domain Compressive Channel Estimation for Frequency-Selective Hybrid Millimeter Wave MIMO Systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 5, pp. 2946–2960, 2018.

[11] S. Roger, M. Cobos, C. Botella-Mascarell, and G. Fodor, "Fast Channel Estimation in the Transformed Spatial Domain for Analog Millimeter Wave Systems," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2021.

[12] X. Lin, S. Wu, C. Jiang, L. Kuang, J. Yan, and L. Hanzo, "Estimation of Broadband Multiuser Millimeter Wave Massive MIMO-OFDM Channels by Exploiting Their Sparse Structure," *IEEE Transactions on Wireless Communications*, vol. 17, no. 6, pp. 3959–3973, 2018.

[13] A. Kaushik, E. Vlachos, J. Thompson, and A. Perelli, "Efficient Channel Estimation in Millimeter Wave Hybrid MIMO Systems with Low Resolution ADCs," in *2018 26th European Signal Processing Conference (EUSIPCO)*, 2018, pp. 1825–1829.

[14] 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, 2018.

[15] M. Bayati and A. Montanari, "The Dynamics of Message Passing on Dense Graphs, with Applications to Compressed Sensing," *IEEE Transactions on Information Theory*, vol. 57, no. 2, pp. 764–785, 2011.

[16] 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, 2014.

[17] A. Maleki, "Approximate Message Passing Algorithm for Compressed Sensing," Ph.D. dissertation, Stanford University, 2010.

[18] D. L. Donoho, A. Maleki, and A. Montanari, "Message Passing Algorithms for Compressed Sensing: I. Motivation and Construction," in *2010 IEEE Information Theory Workshop on Information Theory (ITW 2010, Cairo)*, Jan 2010, pp. 1–5.

[19] N. Goertz and S. Birgmeier, "Sparse Measurement Matrices for Compressed-Sensing Recovery by Bayesian Approximate Message Passing," in *WSA 2020; 24th International ITG Workshop on Smart Antennas*, 2020, pp. 1–6.

[20] M. Mussbah, S. Schwarz, and M. Rupp, "Beam Selection-Based Hybrid Precoding," in *2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*, 2021, pp. 1–6.

[21] M. Mussbah, "Single-Pixel Imaging for Capturing the Shape of Three-Dimensional Objects," Master's thesis, TU Wien, 2020.

[22] C. Ueberhuber, *Numerical Computation 1: Methods, Software, and Analysis*. Springer Berlin Heidelberg, 2012. [Online]. Available: https://books.google.at/books?id=giH7CAAAQBAJ

[23] L. Liang, W. Xu, and X. Dong, "Low-Complexity Hybrid Precoding in Massive Multiuser MIMO Systems," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 653–656, 2014.