Data Rate Bound for mmWave Hybrid Beamforming Systems with Subarrays

Marcin Iwanow[1,2], Nikola Vučić[1], Wolfgang Utschick[2],
Mario H. Castañeda[1], Wen Xu[1], Jian Luo[1]
[1]Huawei Technologies Duesseldorf GmbH, Munich, Germany
Email: {iwanow.marcin, nikola.vucic, mario.castaneda, wen.dr.xu, jianluo}@huawei.com
[2]Associate Institute for Signal Processing, Technische Universität München (TUM), Germany
Email: utschick@tum.de

***Abstract*—In this paper, we consider single-link mmWave transmission. We investigate the usual hybrid beamforming architecture under an additional constraint on the analog precoding stage. Namely, for practical reasons only partial interconnection between the antennas and the RF chains is allowed. We present a novel upper bound for spectral efficiency under this architecture constraint. We evaluate the tightness of the bound by comparing it to the rate achievable with a proposed practical scheme. Finally, we compare the bound with the channel capacity.**

I. Introduction

In recent years, millimeter wave (mmWave) transmission has drawn significant attention as a potential enabler for the fifth generation (5G) cellular networks [1]. This is due to large portions of underutilized spectrum available in those frequency bands. In order to combat the large path loss affecting propagation at mmWave frequencies, large-scale antenna arrays on both sides of the link have to be considered.

Introducing a separate RF chain (including, e.g., an up(down)converter and high-resolution analog-to-digital/digital-to-analog converters) for each antenna is infeasible mainly for power consumption reasons. The literature proposes two main directions that aim on decreasing the power consumption of a massive-antenna device without reducing the antenna gain. One idea is to reduce the resolution of the analog-to-digital converters (ADC's), ultimately even up to 1 bit [2]. In many scenarios, such an approach can guarantee good performance in low and medium SNR range. The second line of research proposes to introduce an analog processing stage between the antennas and the RF chains. The analog stage is a network of phase shifters, combiners and splitters that, at the receiver, reduces the dimension of the received signal to the number of available RF chains. The operation at the transmitter is reverse. In the baseband, full digital processing is available. In the literature, such system is often labeled as hybrid digital-analog beamforming [3]–[5].

In recent works [6]–[9] it has been questioned, whether the usual assumption of full interconnection between the antennas and the RF chains (*fully interconnected hybrid beamforming* (FI-HBF) architecture) is realizable in practice. For example, the analog stage of a hybrid transmitter consisting of 128 antennas and 8 RF chains would require 1024 phase shifters, 128 8-way adders and 8 128-way splitters. The design of such a complex circuit operating reliably on mmWave frequencies is an extreme hardware challenge.

An alternative is the *partially interconnected* (PI-HBF) architecture. It assumes the connection of each antenna only to a subset of all available RF chains. In this paper, we focus on the *subarray-partially interconnected* (SPI) case, when different sets of antennas (subarrays) are connected to *disjoint* sets of RF chains. The motivation for considering this structure is twofold. First, it offers a substantial simplification to the analog circuit. To see this, let's consider the same setup as previously but with 4 subarrays each connected to 2 RF chains. In the SPI-HBF architecture, this requires only 256 phase shifters, 128 2-way adders and 8 16-way splitters. Moreover, such architecture covers also a setup with physically separated subarrays, when full interconnection is not possible.

Our contribution to the existing literature (i.a., [6]–[9]) on SPI-HBF can be summarized as follows:

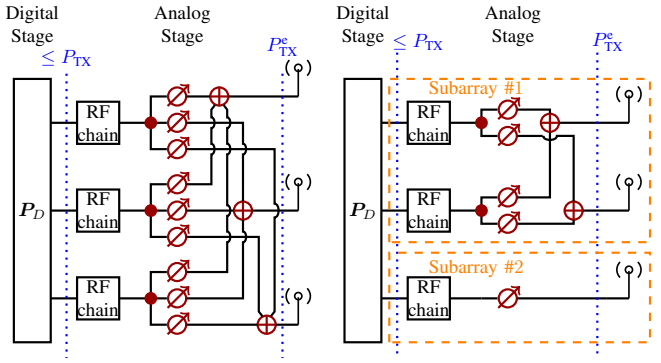
- Fundamental upper bound for the rate is derived. It is based on an analogy with a multiuser scenario and holds for uniform power allocation.
- We propose a practical linear precoding algorithm performing close to the theoretical limit.
- We adopt a passive analog domain network and define the power constraint accordingly.
- The results do not restrict to 1 RF chain per subarray.

A. Notation

We use lower case boldface characters to denote vectors and capital boldface characters to denote matrices. We denote $N \times N$ identity matrix with $\boldsymbol{I}_N$. $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ denote transpose and conjugate transpose. We use $\|\cdot\|_{\mathrm{F}}$ for the Frobenius norm. A submatrix including all rows and a subset of columns is denoted with $[\cdot]_{(:,a:b)}$. The expected value of a random variable is denoted with $\mathrm{E}[\cdot]$ and $\boldsymbol{b} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{a}, \boldsymbol{A})$ means that the random variable $\boldsymbol{b}$ follows a circular complex Gaussian distribution with mean $\boldsymbol{a}$ and covariance matrix $\boldsymbol{A}$. $b \sim U[a_1, a_2]$ means that random variable b follows the uniform distribution within $[a_1, a_2]$.

II. System Model

We consider a multiple-input-multiple-output (MIMO) point-to-point setup with the transmitter (TX) and receiver



(a) Exemplary FI-HBF transmitter structure.

(b) Exemplary SPI-HBF transmitter structure.

Fig. 1. Comparison of the (a) fully and (b) partially interconnected hybrid transmitter structure. The structure of the receiver is similar, with exchanged combiners $\oplus$ and signal splitters $\bullet$.

(RX) equipped with $N^{\text{ant}}_{\text{TX}}$ and $N^{\text{ant}}_{\text{RX}}$ antennas, respectively. The frontends of both the TX and RX have hybrid beamforming structure with $N^{\text{RF}}_{\text{TX}}$ and $N^{\text{RF}}_{\text{RX}}$ RF chains, respectively. We allow for partial interconnection within the analog stage. We divide the antenna arrays into $N_{\text{TX,SA}}$ and $N_{\text{RX,SA}}$ subarrays at the TX and RX, respectively, and assign to them disjoint subsets of all the available RF chains.

The signal recovered after processing reads as

$$\hat{\boldsymbol{y}} = \boldsymbol{G}_D^{\text{H}} \boldsymbol{G}_A^{\text{H}} \boldsymbol{H} \boldsymbol{P}_A \boldsymbol{P}_D \boldsymbol{s} + \boldsymbol{G}_D^{\text{H}} \boldsymbol{G}_A^{\text{H}} \boldsymbol{\eta} \tag{1}$$

where $\boldsymbol{G}_D \in \mathbb{C}^{N^{\text{RF}}_{\text{RX}} \times N_{\text{RF}}}$ and $\boldsymbol{P}_D \in \mathbb{C}^{N^{\text{RF}}_{\text{TX}} \times N_{\text{RF}}}$ are the digital combining and precoding matrices, respectively, $\boldsymbol{s} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \boldsymbol{I}_{N_{\text{RF}}})$ is the input signal, $N_{\text{RF}} = \min(N^{\text{RF}}_{\text{RX}}, N^{\text{RF}}_{\text{TX}})$[1], and $\boldsymbol{\eta} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \boldsymbol{R}_{\eta})$ is the Gaussian noise at the receiver's input. $\boldsymbol{G}_A$ and $\boldsymbol{P}_A$ are the analog combining and precoding matrices, respectively. Their structure is constrained as follows

$$|[\boldsymbol{P}_A]_{i,j}| = \begin{cases} \xi^{(k)}_{A,\text{TX}} & \text{iff the } i\text{th antenna and } j\text{th RF chain} \\ & \text{belong to } k\text{th transmit subarray,} \\ 0 & \text{otherwise,} \end{cases}$$
$$|[\boldsymbol{G}_A]_{i,j}| = \begin{cases} \xi^{(k)}_{A,\text{RX}} & \text{iff the } i\text{th antenna and } j\text{th RF chain} \\ & \text{belong to } k\text{th receive subarray,} \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The fixed-magnitude constraint within each subarray reflects the operation of the phase shifter. We shall see later in (3) that the dimensions of the subarray affect this value.

We denote feasible sets, which account for those restrictions as $\mathcal{P}_{\mathcal{A}}$ and $\mathcal{G}_{\mathcal{A}}$. Moreover, we assume (w.l.o.g) that the antennas belonging to the same subarray have subsequent indices. Consequently, the analog precoding/combining matrices $\boldsymbol{P}_A$ and $\boldsymbol{G}_A$ are of block-diagonal structure. For example, the precoding matrix for the SPI-HBF transmitter presented in Fig. 1b reads

$$\boldsymbol{P}_A^{\text{SPI-HBF}} = \begin{bmatrix} \boldsymbol{P} & \boldsymbol{0} \\ \boldsymbol{0}^T & p \end{bmatrix}, \quad \text{where } \boldsymbol{P} \in \mathbb{C}^{2\times 2}, |[\boldsymbol{P}]_{(i,j)}| = \xi^{(1)}_{A,\text{TX}}$$
$$p \in \mathbb{C}, |p| = \xi^{(2)}_{A,\text{TX}}.$$

In this study, we restrict to passive components in the analog stage. We write therefore (omitting the TX and RX subscripts)

$$\xi^{(k)}_A = \left(N^{\text{RF,k}}_{\text{SA}} N^{\text{ant,k}}_{\text{SA}} \right)^{-1/2} \tag{3}$$

where $N^{\text{ant,k}}_{\text{SA}}$ is the number of antennas at the subarray with index k and $N^{\text{RF,k}}_{\text{SA}}$ reflects the number of RF chains assigned to this subarray. The value can be verified by analyzing the scattering matrices of the power splitters/combiners [10].

Most works on hybrid beamforming architecture (cf. [3], [11]) set the power constraint on the compound analog-digital precoding. In our work, we set the power constraint at the power amplifiers right after the digital precoding stage. We argue that in our model the analog phase shifter network includes exclusively passive elements. Moreover, in our work we restrict to uniform power allocation among the streams, and consequently $\|[\boldsymbol{P}_D]_{(:,k)}\|_2^2 = \frac{P_{\text{TX}}}{N_{\text{RF}}}, \forall k \in \{1, \ldots, N_{\text{RF}}\}$. This assumption fits well for limited feedback system architecture [12] or high SNR scenarios. Moreover, in [13] it has been shown that the digital precoder converges to a scaled unitary matrix in systems with hybrid precoding and large number of antennas.

We denote the power budget available at the transmitter with P_{TX} and the emitted power with P^{e}_{TX}. Due to the properties of the signal vector $\boldsymbol{s}$ we relate this powers with the precoding matrices as follows

$$\|\boldsymbol{P}_D\|_{\text{F}}^2 \leq P_{\text{TX}}, \quad P^{\text{e}}_{\text{TX}} = \|\boldsymbol{P}_A \boldsymbol{P}_D\|_{\text{F}}^2. \tag{4}$$

A. Channel Model

In this paper, we assume a geometric narrowband channel model as, e.g., in [3], which reads as

$$\boldsymbol{H} = \sqrt{\frac{N^{\text{ant}}_{\text{RX}} N^{\text{ant}}_{\text{TX}}}{\sum_{l=1}^{N_{\text{cl}}} N^{l}_{\text{path}}}} \sum_{l=1}^{N_{\text{cl}}} \sum_{r=1}^{N^{l}_{\text{path}}} \alpha_{r,l} \boldsymbol{a}_{\text{RX}}(\theta^{\text{RX}}_{r,l}) \boldsymbol{a}^{\text{H}}_{\text{TX}}(\theta^{\text{TX}}_{r,l}). \tag{5}$$

where N_{cl} is the number of clusters, N^{l}_{path} is the number of paths in the l-th cluster, $\alpha_{r,l}$ is the path gain, $\theta^{\text{RX}}_{r,l}$ and $\theta^{\text{TX}}_{r,l}$ are the angles of arrival (AoA) and departure (AoD), respectively. The vectors $\boldsymbol{a}_{\text{RX}} \in \mathbb{C}^{N^{\text{ant}}_{\text{RX}}}$ and $\boldsymbol{a}_{\text{TX}} \in \mathbb{C}^{N^{\text{ant}}_{\text{TX}}}$ are compound antenna array response vectors for the receiver and the transmitter, respectively which contain concatenated array response vectors of all the subarrays. We assume that the subarrays together form a uniform linear array (ULA) with half-wavelength antenna spacing and therefore the array response vectors have the form

$$\boldsymbol{a}_{\text{ULA}}(\theta) = \frac{1}{N_{\text{ant}}} \left[1, e^{j\pi \sin(\theta)}, \ldots, e^{j(N_{\text{ant}}-1)\pi \sin(\theta)} \right]. \tag{6}$$

We assume perfect knowledge of the channel matrix at both the RX and TX, leaving the analysis for imperfect/estimated channel for further work.

[1]In the entire paper, we use N_{RF} to denote the number of streams. It is in order to stress that we assume maximal number of streams, which is limited by the number of RF chains.

III. Bounds on the Achievable Rate

A. The Capacity Constrained by the Number of RF chains

The achievable rate for any hybrid transceiver setup can be upper-bounded by providing the TX and RX with full processing capabilities (among others, optimal power allocation) and leaving only the limitation of the number of RF chains. The capacity constrained by such a limitation reads [14]

$$C_{\text{FD}} = \log_2 \det \left(\boldsymbol{I}_{N_{\text{RX}}^{\text{ant}}} + \boldsymbol{R}_\eta^{-1} \boldsymbol{H} \boldsymbol{V} \boldsymbol{\Phi} \boldsymbol{V}^{\text{H}} \boldsymbol{H}^{\text{H}} \right) \tag{7}$$

where $\boldsymbol{V} \in \mathbb{C}^{N_{\text{TX}}^{\text{ant}} \times N_{\text{RF}}}$ contains eigenvectors of $\boldsymbol{H}^{\text{H}} \boldsymbol{R}_\eta^{-1} \boldsymbol{H}$ corresponding to the N_{RF} largest eigenvalues and $\boldsymbol{\Phi} \in {\mathbb{R}_{++}}^{N_{\text{RF}} \times N_{\text{RF}}}$ is a diagonal matrix with entries corresponding to optimal power allocation along the streams and $\text{tr}(\boldsymbol{\Phi}) = P_{\text{TX}}^{\text{e}}$.

We observe that for any applied analog precoding/combining and uniform power allocation, the digital precoding matrix $\boldsymbol{P}_D$ that maximizes the capacity is a scaled unitary matrix $\boldsymbol{P}_D \boldsymbol{P}_D^{\text{H}} = \frac{P_{\text{TX}}}{N_{\text{TX}}^{\text{RF}}} \boldsymbol{I}_{N_{\text{TX}}^{\text{RF}}}$. It stems from the SVD decomposition of the effective channel, which includes the analog processing stages. Then, P_{TX}^{e} is expressed in a closed form and reads

$$\begin{aligned} P_{\text{TX}}^{\text{e}} &= \text{E}\left[\|\boldsymbol{P}_A \boldsymbol{P}_D \boldsymbol{s}\|_2^2 \right] = \frac{P_{\text{TX}}}{N_{\text{TX}}^{\text{RF}}} \text{tr}(\boldsymbol{P}_A^{\text{H}} \boldsymbol{P}_A) = \\ &= \frac{P_{\text{TX}}}{N_{\text{TX}}^{\text{RF}}} \sum_{k=1}^{N_{\text{TX,SA}}} {\xi_{A,\text{TX}}^{(k)}}^2 N_{\text{TX,SA}}^{\text{ant,k}} N_{\text{TX,SA}}^{\text{RF,k}} = \frac{P_{\text{TX}}}{N_{\text{TX}}^{\text{RF}}} N_{\text{TX,SA}} \leq P_{\text{TX}} \end{aligned} \tag{8}$$

We note that the power constraint is dependent on the subarray configuration at the TX.

B. Novel Achievable Rate Bound

As mentioned in the previous subsection, maximizing the rate of the link requires a scaled unitary digital precoding matrix. Consequently, the optimal digital precoder only scales the covariance matrix of the input data and therefore can be discarded in the analysis. We further consider the following link

$$\boldsymbol{y} = \boldsymbol{H} \boldsymbol{P}_A \boldsymbol{s}' + \boldsymbol{\eta} \tag{9}$$

where $\boldsymbol{s}' = \boldsymbol{P}_D \boldsymbol{s}$. The matrix $\boldsymbol{P}_A$ is block-diagonal with blocks $\boldsymbol{P}_A^1, \ldots, \boldsymbol{P}_A^{N_{\text{TX,SA}}}$ of dimensions $\boldsymbol{P}_A^k \in \mathbb{C}^{N_{\text{TX,SA}}^{\text{ant,k}} \times N_{\text{TX,SA}}^{\text{RF,k}}}$. This structure of the precoding matrix allows for building an analogy to the multiple access channel (MAC) setup. To make this clearer, we rewrite (9) as follows

$$\boldsymbol{y} = \sum_{k=1}^{N_{\text{TX,SA}}} \boldsymbol{H}_k \boldsymbol{P}_A^k \boldsymbol{s}_k' + \boldsymbol{\eta} \tag{10}$$

where $\boldsymbol{H}_k \in \mathbb{C}^{N_{\text{RX}}^{\text{ant}} \times N_{\text{TX,SA}}^{\text{ant,k}}}$, $\boldsymbol{H} = \left[\boldsymbol{H}_1, \ldots, \boldsymbol{H}_{N_{\text{TX,SA}}} \right]$ and $\boldsymbol{s}_k' \in \mathbb{C}^{N_{\text{TX,SA}}^{\text{RF,k}}}$, $\boldsymbol{s}' = [{\boldsymbol{s}_1'}^{\text{T}}, \ldots {\boldsymbol{s}_{N_{\text{TX,SA}}}'}^{\text{T}}]^{\text{T}}$.

The rate of this link is upper bounded by assuming perfect (nearest neighbor) reception. In such case, the rate is equal to the sum rate of the MAC setup with subarrays acting as users. We further note that joint coding among the subarrays could not improve the rate. Although joint coding can expand the MAC rate region, it does not affect the sum rate [15].

Algorithm 1 Precoder-based iterative waterfilling with individual power constraints

Require: $\boldsymbol{H}_k$, $\boldsymbol{R}_\eta$, P_{TX}^k

$\boldsymbol{P}_A^k \leftarrow [\boldsymbol{I}]_{(:,1:N_{\text{TX,SA}})}$, $\boldsymbol{Q}_k \leftarrow \boldsymbol{P}_A^k {\boldsymbol{P}_A^k}^{\text{H}} \quad \forall k$

$\boldsymbol{\Sigma} \leftarrow \boldsymbol{R}_\eta + \frac{P_{\text{TX}}}{N_{\text{TX}}^{\text{RF}}} \sum_k^{N_{\text{TX,SA}}} \boldsymbol{H}_k \boldsymbol{Q}_k \boldsymbol{H}_k^{\text{H}}$

repeat

  for $k = 1, \ldots, N_{\text{TX,SA}}$ **do**

    Update noise+interference cov. matrix at kth SA

    $\boldsymbol{N}_k = \boldsymbol{\Sigma} - \boldsymbol{H}_k \boldsymbol{Q}_k \boldsymbol{H}_k^{\text{H}}$

    Eigenvalue decomposition of the effective channel

    $\boldsymbol{H}_k^{\text{H}} \boldsymbol{N}_k^{-1} \boldsymbol{H}_k = \boldsymbol{W}_k \boldsymbol{\Lambda}_k \boldsymbol{W}_k^{\text{H}}$

    Waterfilling (μ_k - water level)

    $\boldsymbol{\Phi}_k = \max(\mu_k \boldsymbol{I} - \boldsymbol{\Lambda}_k^{-1}, \boldsymbol{0})$ s.t. $\text{tr}(\boldsymbol{\Phi}_k) = P_{\text{TX}}^k$

    Update precoders and TX/RX covariance matrices

    $\boldsymbol{P}_A^k = \boldsymbol{W}_k [\boldsymbol{\Phi}_k]_{(:,1:N_{\text{TX,SA}}^{\text{RF,k}})}$, $\boldsymbol{Q}_k = \boldsymbol{P}_A^k {\boldsymbol{P}_A^k}^{\text{H}}$

    $\boldsymbol{\Sigma} = \boldsymbol{N}_k + \frac{P_{\text{TX}}}{N_{\text{TX}}^{\text{RF}}} \sum_k^{N_{\text{TX,SA}}} \boldsymbol{H}_k \boldsymbol{Q}_k \boldsymbol{H}_k^{\text{H}}$

until convergence

return $\boldsymbol{P}_A^k \quad \forall k \in \{1, \ldots, N_{\text{TX,SA}}\}$

The MAC sum rate written as a function of the precoding matrices $\boldsymbol{P}_A^k$ reads

$$\begin{aligned} R\left(\boldsymbol{P}_A^1, \ldots, \boldsymbol{P}_A^{N_{\text{TX,SA}}} \right) = &- \log_2 \det(\boldsymbol{R}_\eta) + \\ &\log_2 \det \left(\boldsymbol{R}_\eta + \frac{P_{\text{TX}}}{N_{\text{TX}}^{\text{RF}}} \sum_{k=1}^{N_{\text{TX,SA}}} \boldsymbol{H}_k \boldsymbol{P}_A^k \boldsymbol{P}_A^{k,\text{H}} \boldsymbol{H}_k^{\text{H}} \right) \end{aligned} \tag{11}$$

We further relax the constant-magnitude constraint $|[\boldsymbol{P}_A^k]_{(i,j)}| = \xi_{A,\text{TX}}^{(k)}$ to a per subarray power constraint which reads $\text{tr}(\boldsymbol{P}_A^k {\boldsymbol{P}_A^k}^{\text{H}}) \leq P_{\text{TX}}^k$ where $P_{\text{TX}}^k = {\xi_{A,\text{TX}}^{(k)}}^2 N_{\text{TX,SA}}^{\text{ant,k}} N_{\text{TX,SA}}^{\text{RF,k}}$, and write the MAC sum-rate maximization problem as

$$\begin{aligned} C_{\text{subarr}}^{\text{MAC}} = &\max_{\boldsymbol{P}_A^1, \ldots, \boldsymbol{P}_A^{N_{\text{TX,SA}}}} R(\boldsymbol{P}_A^1, \ldots, \boldsymbol{P}_A^{N_{\text{TX,SA}}}) \\ &\text{s.t. } \text{tr}(\boldsymbol{P}_A^k {\boldsymbol{P}_A^k}^{\text{H}}) \leq P_{\text{TX}}^k \quad \forall k. \end{aligned} \tag{12}$$

The classical work on iterative waterfilling involves optimization of the MAC sum rate with respect to the covariance matrices [16], [17]. In [18], a precoder-based solution involving the sum power constraint has been presented. Inspired by it, we present in Algorithm 1 a simple extension to the classical iterative waterfilling algorithm, such that it iteratively updates the precoders and therefore is perfectly suited to solve (12).

IV. Practical Precoding Scheme

In this section, we design a precoding/combining strategy for the SPI-HBF architecture. To this end, we present a way to extend the existing algorithms for FI-HBF, which base on decomposing the optimal unconstrained precoding/combining

Algorithm 2 BCD-based solution to (14)

Require: $\boldsymbol{P}^{k,\star}$

1: $\boldsymbol{P}_A^k \in \mathbb{C}^{N_{\text{TX,SA}}^{\text{ant,}} \times N_{\text{TX,SA}}^{\text{RF,}}} \leftarrow$ Arbitrary
2: **repeat**
3: $\quad \boldsymbol{P}_D^k \leftarrow \left(\boldsymbol{P}_A^{k\,\text{H}} \boldsymbol{P}_A^k\right)^{-1} \boldsymbol{P}_A^{k\,\text{H}} \boldsymbol{P}^{k,\star}$
$\quad$ projection into $\mathcal{P}_{\mathcal{A}}$: $[\Pi_{\mathcal{P}_{\mathcal{A}}}[\boldsymbol{A}]]_{(i,j)} = \frac{[\boldsymbol{A}]_{(i,j)}}{|[\boldsymbol{A}]_{(i,j)}|} \xi_{A,\text{TX}}^{(k)}$
4: $\quad \boldsymbol{P}_A^k \leftarrow \Pi_{\mathcal{P}_{\mathcal{A}}}\left[\boldsymbol{P}^{k,\star} \boldsymbol{P}_D^{k\,\text{H}} \left(\boldsymbol{P}_D^k \boldsymbol{P}_D^{k\,\text{H}}\right)\right]$
5: **until** convergence
6: **return** $\boldsymbol{P}_A^k, \boldsymbol{P}_D^k$

matrices into a product of the analog and digital precoding/combining matrices. Below we describe the procedure for the TX, as it differs to the RX decomposition only w.r.t. the power constraint [4].

Like in [3], [4], [11] we aim on minimizing the Frobenius norm of the decomposition error, namely

$$(\boldsymbol{P}_A^\star, \boldsymbol{P}_D^\star) = \underset{\boldsymbol{P}_A, \boldsymbol{P}_D}{\arg\min} \quad \|\boldsymbol{P}^\star - \boldsymbol{P}_A \boldsymbol{P}_D\|_{\text{F}} \quad \text{s.t.} \quad \|\boldsymbol{P}_D\|_{\text{F}}^2 \leq P_{\text{TX}}, \quad \boldsymbol{P}_A \in \mathcal{P}_{\mathcal{A}} \tag{13}$$

where $\boldsymbol{P}^\star = [\boldsymbol{V}]_{(:,1:N_{\text{RF}})}$ is the optimal unconstrained precoding matrix. The matrix $[\boldsymbol{V}]_{(:,1:N_{\text{RF}})}$ contains eigenvectors corresponding to the N_{RF} largest eigenvalues of $\boldsymbol{H}^{\text{H}} \boldsymbol{R}_\eta^{-1} \boldsymbol{H}$.

Next, we exploit the block-diagonal structure of $\boldsymbol{P}_A$ and write (13) equivalently as

$$(\boldsymbol{P}_A^\star, \boldsymbol{P}_D^\star) = \underset{\boldsymbol{P}_A, \boldsymbol{P}_D}{\arg\min} \quad \sum_{k=1}^{N_{\text{TX,SA}}} \|\boldsymbol{P}^{k,\star} - \boldsymbol{P}_A^k \boldsymbol{P}_D^k\|_{\text{F}} \quad \text{s.t.} \quad \|\boldsymbol{P}_D\|_{\text{F}}^2 \leq P_{\text{TX}}, \quad |[\boldsymbol{P}_A^k]_{i,j}| = \xi_A^k \quad \forall_{i,j} \tag{14}$$

where $\boldsymbol{P}_A^k$ are the diagonal blocks of $\boldsymbol{P}_A$ and

$$\boldsymbol{P}_D = \left[\boldsymbol{P}_D^{1\,\text{T}}, \ldots, \boldsymbol{P}_D^{N_{\text{TX,SA}}\,\text{T}}\right]^{\text{T}} \text{ with } \boldsymbol{P}_D^k \in \mathbb{C}^{N_{\text{TX,SA}}^{\text{RF,k}} \times N_{\text{RF}}}.$$

and the matrix $\boldsymbol{P}_D^\star$ is defined analogously.

The optimization variables in the individual terms of the objective function in (14) can be solved independently by means of, e.g., the sparse recovery *orthogonal matching pursuit* OMP algorithm or the modified *block coordinate descent* (BCD) method [3], [4], [8]. We present the BCD solution in Algorithm 2.

In order to account for the uniform power allocation which is assumed, we perform an additional step at the end. We design the precoding/combining matrices of the digital stages based on the effective channel matrix $\boldsymbol{H}_{\text{eff}} = \boldsymbol{G}_A^{\text{H}} \boldsymbol{H} \boldsymbol{P}_A$. They read

$$\boldsymbol{P}_D^\star = [\boldsymbol{V}_{\text{eff}}]_{(:,1:N_{\text{RF}})}, \quad \boldsymbol{G}_D^\star = \boldsymbol{P}_D^{\star,\text{H}} \boldsymbol{H}_{\text{eff}}^{\text{H}} \boldsymbol{R}_\eta^{-1} \tag{15}$$

where $[\boldsymbol{V}_{\text{eff}}]_{(:,1:N_{\text{RF}})}$ are the eigenvectors corresponding to the N_{RF} largest eigenvalues of $\boldsymbol{H}_{\text{eff}}^{\text{H}} \boldsymbol{R}_\eta^{-1} \boldsymbol{H}_{\text{eff}}$. The rate achievable with this scheme reads

$$R_{\text{decomp}} = \log_2 \det\left(\boldsymbol{I}_{N_{\text{RX}}^{\text{ant}}} + \boldsymbol{R}_\eta^{A\,-1} \boldsymbol{H}_{\text{eff}} \boldsymbol{H}_{\text{eff}}^{\text{H}}\right) \tag{16}$$

with $\boldsymbol{R}_\eta^A = \boldsymbol{G}_A^{\star\,\text{H}} \boldsymbol{R}_\eta \boldsymbol{G}_A^\star$ and $\boldsymbol{H}_{\text{eff}} = \boldsymbol{G}_D^{\star\,\text{H}} \boldsymbol{G}_A^{\star\,\text{H}} \boldsymbol{H} \boldsymbol{P}_A^\star \boldsymbol{P}_D^\star$.

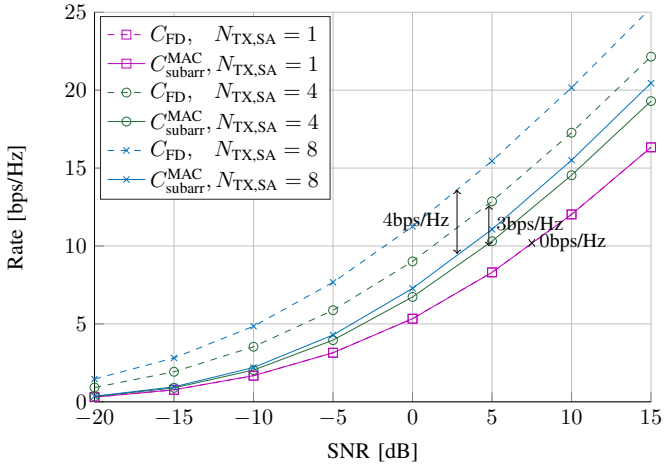


Fig. 2. Comparison of the constrained capacity C_{FD} and new upper bound $C_{\text{subarr}}^{\text{MAC}}$ for different number of subarrays at the TX. $N_{\text{TX}}^{\text{ant}} = 64$, $N_{\text{RX}}^{\text{ant}} = 16$, $N_{\text{TX}}^{\text{RF}} = N_{\text{RX}}^{\text{RF}} = 8$.

V. Numerical Results

In the simulations, we compare the novel bound with the constrained capacity C_{FD}. Also, we evaluate the tightness of the bound by comparing it with the rate achievable with the practical scheme proposed in Section IV.

We assume $N_{\text{cl}} = 3$ clusters with $N_{\text{path}}^l = 10$ paths each. The central cluster angles drawn at random from the uniform distribution $U[0, 2\pi]$. The angles of arrival/departures within the cluster are drawn at random from a Laplacian distribution with variance of $8°$ around the central cluster angle. The path gains follow normal complex Gaussian distribution $\alpha_{r,l} \sim \mathcal{N}_{\mathbb{C}}(0, 1)$.

The TX and RX are equipped wit $N_{\text{TX}}^{\text{ant}} = 64$ and $N_{\text{RX}}^{\text{ant}} = 16$ antennas, respectively. Both the TX and RX have $N_{\text{RF}} = 8$ RF chains unless stated differently. The antennas and RF chains are equally distributed among the subarrays.

The first set of results depicted in Fig. 2 compares $C_{\text{subarr}}^{\text{MAC}}$ with the constrained capacity C_{FD}. We observe that the bounds are equal for FI-HBF and the $C_{\text{subarr}}^{\text{MAC}}$ is tighter than C_{FD} as the number of subarrays at the transmitter rises. This is expected as the bound is designed such that it takes into account the SPI-HBF structure.

In Fig. 3 we investigate the relation of the achievable rate R_{decomp} and $C_{\text{subarr}}^{\text{MAC}}$. We observe that the bound is tight especially for many subarrays at the TX. Moreover we see that for the setup considered, the number of subarrays at the RX has minimal influence on the performance.

A different angle on investigation of the bound tightness is provided in Fig. 4. There, we observe the separation between $C_{\text{subarr}}^{\text{MAC}}$ and R_{decomp} at SNR = 0 dB for different numbers of RF chains available. We conclude two dependencies from the plot. First, if more RF chains are available, the impact of the subarray structure at the RX is lower. Second is consistent with the observations based on Fig. 3. Namely, the bound becomes tighter as the number of subarrays at the TX increases.

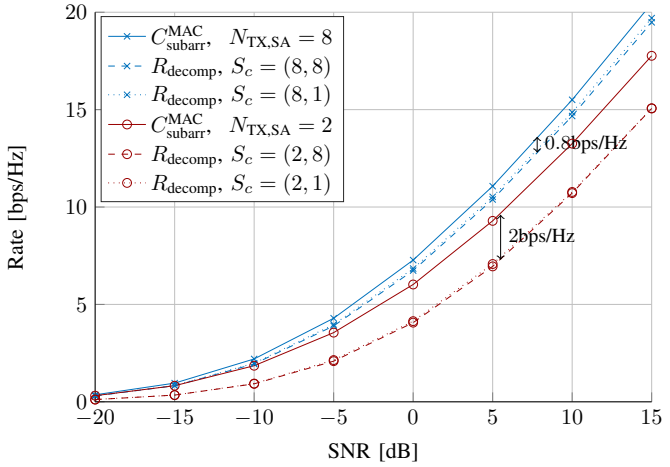


Fig. 3. Comparison of the derived upper bound with the rate achievable by the scheme from Section IV. $S_c = (m, n)$ is the subarray configuration and stands for $N_{\text{TX,SA}} = m, N_{\text{RX,SA}} = n$. $N^{\text{ant}}_{\text{TX}} = 64$, $N^{\text{ant}}_{\text{RX}} = 16$, $N^{\text{RF}}_{\text{TX}} = N^{\text{RF}}_{\text{RX}} = 8$.

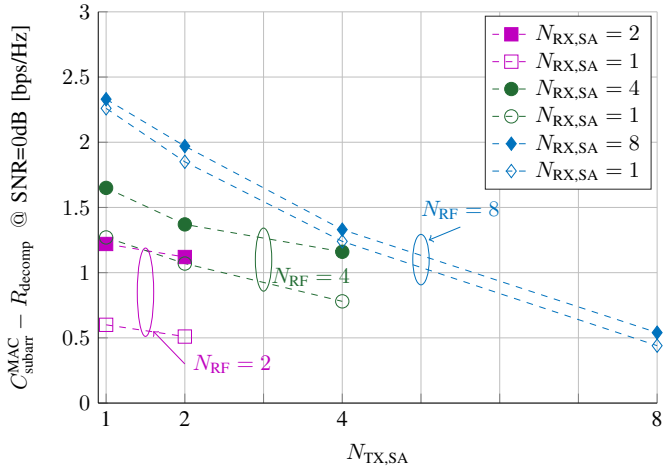


Fig. 4. Distance between the upper bound $C^{\text{MAC}}_{\text{subarr}}$ and the achievable rate R_{decomp} at SNR= 0 dB for transceivers with different number of RF chains.

VI. Conclusions

In the paper, a novel upper bound for the rate achievable with SPI-HBF transceiver structure has been presented. The bound appears to be very tight if many subarrays are present at the transmitter. Moreover, we showed that the definition of the power constraint can have significant influence on the transceiver design. For the approach that we followed, i.e., assuming a power constraint after the digital precoding and allocating equal power to the streams, less subarrays at the TX result in decreased mean emitted power. Consequently, the FI-HBF does not necessarily provide the best performance and implementing SPI-HBF can be advantageous.

We also stress the application of the multiuser perspective, where the subarrays are acting as users. The results encourage to consider this perspective in further work on SPI-HBF.

Acknowledgment

This work has been supported in part by the EU Horizon 2020 5GPPP project 5G-Xhaul. The authors would also like to thank Ronald Böhnke (Huawei Technologies Duesseldorf) and Michael Joham (Technische Universität München) for fruitful discussions.

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*, pp. 335–349, May 2013.

[2] J. Mo and R. W. Heath, "Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information," *IEEE Transactions on Signal Processing*, vol. 63, no. 20, pp. 5498–5512, Oct. 2015.

[3] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[4] M. Iwanow, N. Vucic, M. Castaneda, J. Luo, W. Xu, and W. Utschick, "Some aspects on hybrid wideband transceiver design for mmWave communication systems," in *ITG Workshop on Smart Antennas (WSA)*, Munich, Mar. 2016.

[5] A. Adhikary, E. Al Safadi, M. Samimi, R. Wang, G. Caire, T. Rappaport, and A. Molisch, "Joint spatial division and multiplexing for mm-Wave channels," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1239–1255, June 2014.

[6] X. Gao, L. Dai, S. Han, C. L. I, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmwave MIMO systems with large antenna arrays," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 998–1009, Apr. 2016.

[7] O. E. Ayach, R. W. Heath, S. Rajagopal, and Z. Pi, "Multimode precoding in millimeter wave MIMO transmitters with multiple antenna sub-arrays," in *Proc. IEEE Global Communications Conf. (GLOBECOM)*, Dec. 2013, pp. 3476–3480.

[8] X. Yu, J. C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 485–500, Apr. 2016.

[9] S. He, C. Qi, Y. Wu, and Y. Huang, "Energy-efficient transceiver design for hybrid sub-array architecture MIMO systems," *IEEE Access*, vol. 4, pp. 9895–9905, 2016.

[10] D. M. Pozar, *Microwave engineering*. John Wiley & Sons, 2009.

[11] H. Ghauch, T. Kim, M. Bengtsson, and M. Skoglund, "Subspace estimation and decomposition for large millimeter-wave MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 528–542, Apr. 2016.

[12] D. J. Love and R. W. Heath, "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Transactions on Information Theory*, vol. 51, no. 8, pp. 2967–2976, Aug. 2005.

[13] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 501–513, Apr. 2016.

[14] D. P. Palomar, "A unified framework for communications through MIMO channels," Ph.D. dissertation, Universitat Politecnica de Catalunya, May 2003.

[15] F. Willems, "The discrete memoryless multiple access channel with partially cooperating encoders (corresp.)," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 441–445, May 1983.

[16] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for Gaussian vector multiple-access channels," *IEEE Transactions on Information Theory*, vol. 50, no. 1, p. 145152, Jan 2004.

[17] R. Böhnke, V. Kühn, and K. Kammeyer, "Fast sum rate maximization for the downlink of MIMO-OFDM systems," in *Canadian Workshop on Information Theory (CWIT)*, Montreal, Canada, 2005.

[18] R. Hunger, D. A. Schmidt, and W. Utschick, "Sum-capacity and MMSE for the MIMO broadcast channel without eigenvalue decompositions," in *2007 IEEE International Symposium on Information Theory*, Jun. 2007, pp. 776–780.