Receive Antenna Selection and Hybrid Precoding for Receive Spatial Modulation in Massive MIMO Systems

Ahmed Raafat, Adrian Agustin and Josep Vidal
Dept. of Signal Theory and Communications, Universitat Politecnica de Catalunya (UPC), Barcelona, Spain.
Email: {ahmed.raafat, adrian.agustin, josep.vidal}@upc.edu

***Abstract*—Recently, a receive spatial modulation (RSM) for massive multiple-input-multiple-output operating in millimeter wave (mmWave) was introduced with the purpose of simplifying user terminal circuit by employing only one radio-frequency chain and attaining high spectral efficiency by exploiting the receive spatial dimension. However, when RSM is applied in a mmWave channel, it demands a challenging receive antenna selection (RAS) procedure. On the other hand, the power consumption at the transmitter side is high when a full digital (FD) precoder is envisioned. We consider the joint problem of RAS and precoder designs based low complexity hybrid architecture. For the sake of simplicity, we divide this problem into two subproblems. First, we design the RAS assuming FD precoder, and then, we design the hybrid precoder. We propose two novel and efficient RAS methods. First, we formulate the RAS as non-convex optimization problem. Then, we convert it into a convex optimization problem by introducing novel lower bounds and relaxing non-convex constraints. Second, we provide sequential algorithms that approach the optimal selection where we (add/remove) one (good/poor) antenna per iteration. We propose novel zero forcing hybrid precoder based convex optimization that maximizes the received power. We prove that the proposed precoder is optimal when the channel is highly spatially sparse. The proposed designs have been compared with the best known methods in terms of average mutual information and energy efficiency showing significant improvements.**

I. Introduction

The vast available spectrum of millimeter wave (mmWave) frequency band can significantly enhance the achievable rates of future cellular systems [1]. However, propagation losses in these bands are huge, an effect that can be compensated by the beamforming gains obtained if packing a large number of antennas at the transceivers. Cost and power consumption of fully digital (FD) multiple-input-multiple-output (MIMO) transceiver highly increase at mmWave band [1]. Hence, FD massive MIMO transceiver design becomes challenging.

Spatial transmission is a powerful tool that can be exploited to simplify MIMO transceiver and attain high data rates. Receive spatial modulation (RSM) schemes have been developed with the aim of improving MIMO spectral efficiency by exploiting the receive spatial dimension as an extra information source [2]-[3]-[4]. These schemes have been introduced for sub-6 GHz considering rich multipath environment and FD MIMO transceiver and suffer from performance degradation and high complexity transceiver when applied to mmWave communications. Low complexity RSM MIMO transceiver has been reported in [5] for indoor line-of-sight mmWave communication. Recently, simple RSM MIMO transceiver and novel detection method have been introduced in [6] for outdoor narrowband mmWave communication. Nevertheless, the system in [6] relies on computationally complex receive antenna subset selection (RAS) algorithm and power hungry FD base station (BS) as illustrated in Fig. 1.

Inspired by fast algorithms [7] and by convex optimization [8], several RAS techniques have been studied to maximize MIMO channel spectral efficiency. However, these methods are not alleviating the problems associated to zero forcing (ZF) precoding in [6]. ZF hybrid precoding has been studied in [9] to simplify MIMO BS, whereby signal processing is divided among digital processing at baseband and analog processing at passband. In [10], the authors developed ZF hybrid precoder that can achieve data rates higher to those in [9]. However, the design in [10] is computationally complex and suboptimal.

Considering the RSM system in Fig. 1, we study the RAS problem and ZF hybrid precoder design. The major contributions of this paper are as follows:

- We develop novel, fast and efficient RAS methods.
- We derive closed form expression for mutual information of RSM system in Fig. 1.
- We determine the optimal number of active receive antennas (ARA) by maximizing the mutual information using fast algorithm.
- We develop novel ZF hybrid precoder that has the same performance as the FD precoder when channel is highly spatially sparse and outperforms the design in [10] in achievable rates, energy efficiency and complexity.

We use the following notation through this paper: $(.)^T$ and $(.)^H$ are transpose and conjugate transpose, respectively. $\|\mathbf{X}\|_F$ and $\mathrm{Tr}\{\mathbf{X}\}$ denote Frobenius norm and trace of matrix $\mathbf{X}$, respectively. $|x|$ and $\mathrm{Arg}(x)$ are magnitude and phase of x, respectively. $\mathbf{X}(k,:)$ and $\mathbf{X}_k$ denote k^{th} row and k^{th} diagonal entry of matrix $\mathbf{X}$, respectively. $\mathbf{X}(n,m)$ is entry in the n^{th} row and the m^{th} column of matrix $\mathbf{X}$. $\mathbf{V}_N\{\mathbf{X}\}$ denotes the

The research leading to these results has been partially funded by the 5Gwireless project within the framework of H2020 Marie Skłodowska-Curie innovative training networks (ITNs), the project 5G&B RUNNER-UPC (TEC2016-77148-C2-1-R (AEI/FEDER, UE)) and the Catalan Government (2017 SGR 578-AGAUR).

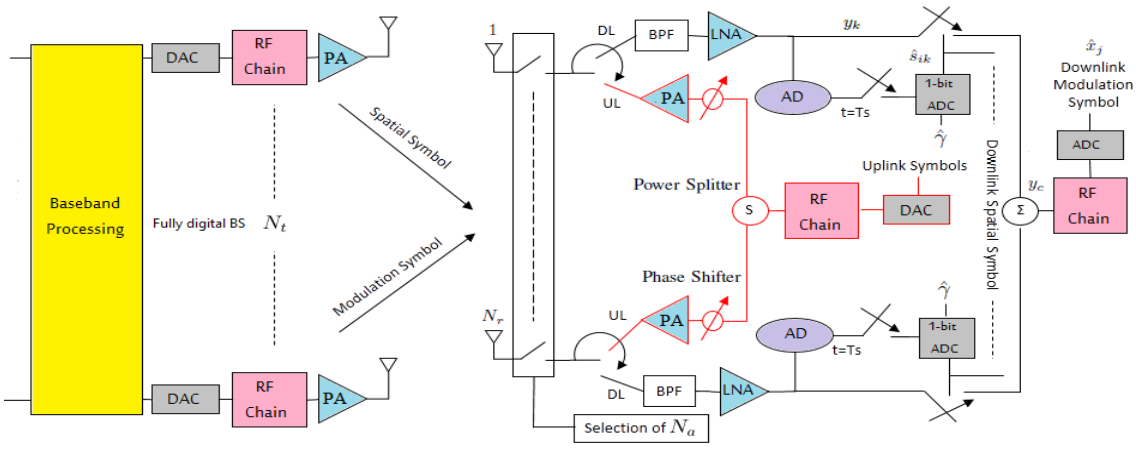

Fig. 1. RSM Massive MIMO transceiver architecture introduced in [6], downlink (black) and uplink (red).

matrix contains the largest N eigenvectors of matrix $\mathbf{X}$.

II. System and Channel Models

We consider the downlink (DL) of single user massive MIMO systems operating in mmWave outdoor narrowband propagation environment where the BS and the user terminal (UT) are equipped with N_t and N_r antennas respectively. Although we consider narrowband signals, the proposed algorithms can be extended to the wideband transmission by applying orthogonal frequency-division multiplexing and this is a future work topic. Since outdoor mmWave propagation suffers from acute path loss, the corresponding channel is limited by few scattering clusters and becomes spatially sparse. Therefore, we adopt the widely used outdoor narrowband channel model in [11] to design and evaluate the proposed system. The channel matrix in this model can be expressed as

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{\xi}} \sum_{i=1}^{L} g_i \mathbf{v}_r\left(\theta_i\right) \mathbf{v}_t\left(\phi_i\right)^H \tag{1}$$

where $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$, ξ is the path loss, L is number of scattering paths, $g_i \in \mathcal{CN}(0,1)$ is independent and identically distributed (i.i.d.) i^{th} path complex gain, $(\phi_i \in [0, 2\pi], \theta_i \in [0, 2\pi])$ are azimuth and elevation angles of departures and arrivals, $\mathbf{v}_t\left(\phi_i\right), \mathbf{v}_r\left(\theta_i\right)$ are the transmit and receive array response vectors. We consider an N-element uniform linear whose response vector can be expressed as

$$\mathbf{v}(\phi) = \frac{1}{\sqrt{N}}\left[1, e^{jkd\sin(\phi)}, ..., e^{j(N-1)kd\sin(\phi)}\right]^T \tag{2}$$

where $k = \frac{2\pi}{\lambda}$ and d is the inter-element spacing.

III. RSM for Massive MIMO Systems

In [6], a low complexity RSM massive MIMO transceiver architecture has been developed based on a FD transmitter and energy efficient receiver circuitry. To the best of our knowledge, architecture in Fig. 1 is the first that combines low complexity RSM massive MIMO transceiver with RAS for spatially sparse channels [6]. In the following subsections, we summarize how the RSM displayed in Fig. 1 is working, its benefits, challenges and the proposed improvements.

A. UT circuit energy efficiency

The UT circuit is designed based on the use of energy efficient devices (amplitude detector (AD) [12], 1-bit analog-to-digital-converter (ADC) and phase shifter [1]) and only one of any power hungry device (high resolution ADC and radio-frequency (RF) chain [1]). The AD can measure amplitude of mmWave RF signal with high sensitivity, negligible power consumption and very high input impedance [12].

B. Transmission protocol

In [6], the authors considered time-division-duplex (TDD) protocol based on DL and uplink (UL) reciprocal environment. The UT does not need the CSI and the BS can acquire the channel knowledge with a low training overhead [6]. At first, the BS acquires the CSI during UL training. Next, the UT estimates the detection threshold $\hat{\gamma}$ (shown in Fig. 1) during DL training using just one pilot symbol [6].

C. Precoding and detection

The sparse nature of mmWave propagation leads to correlation among receive antennas. Therefore, the BS selects the best $(N_a \leq N_r)$ receive antennas to be active and informs the UT about those antennas over a control channel. The received signal vector in Fig. 1 can be expressed as

$$\mathbf{y} = \sqrt{\alpha P}\mathbf{H}_a\mathbf{B}\mathbf{s}_i x_j + \mathbf{n} \tag{3}$$

where $x_j \sim \mathcal{CN}(0,1)$ is the modulation symbol, $\mathbf{s}_i \in \mathbb{R}^{N_a \times 1}$ is a binary spatial symbol conveying N_a data bits, $\mathbf{H}_a \in \mathbb{C}^{N_a \times N_t}$ is the channel between the BS and ARA of the UT, P is average transmit power, $\alpha \approx \left(.5 \times \text{Tr}\left\{\mathbf{B}^H\mathbf{B}\right\}\right)^{-1}$ is a normalization factor that fixes the average transmit power, $\mathbf{n} \in \mathbb{C}^{N_a \times 1}$ noise vector whose entries are i.i.d. $\mathcal{CN}\left(0, \sigma^2\right)$ and $\mathbf{B} \in \mathbb{C}^{N_t \times N_a}$ is the ZF precoder that can be expressed as

$$\mathbf{B} = \mathbf{H}_a^H\left(\mathbf{H}_a\mathbf{H}_a^H\right)^{-1} \tag{4}$$

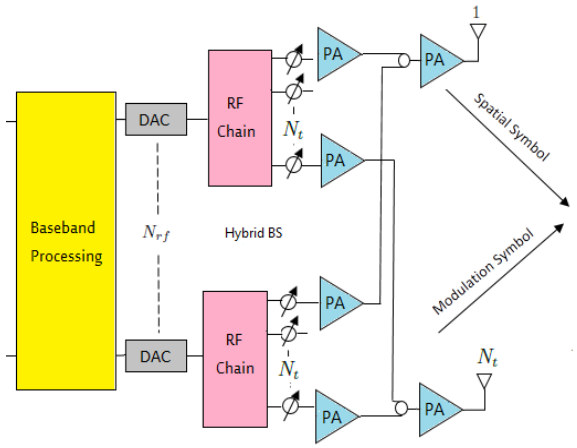

Fig. 2. Fully connected hybrid base station [13].

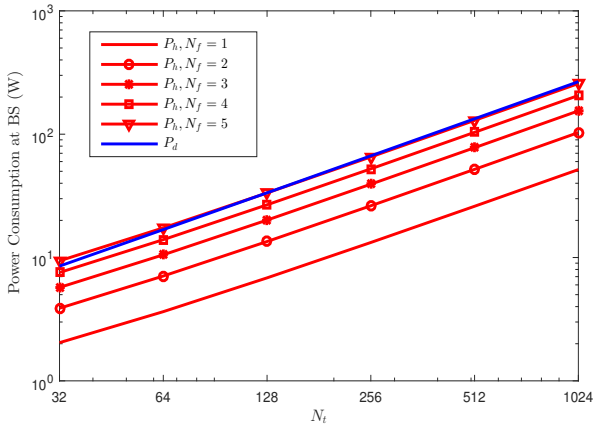

Fig. 3. Hardware power consumption at BS versus N_t at $P_{\text{ref}} = 20$mW.

The received signal at the k^{th} antenna can be expressed as

$$y_k = \sqrt{\alpha P} s_{ik} x_j + n_k \tag{5}$$

where s_{ik} is the k^{th} element of $\mathbf{s}_i$.

RAS is necessary to perform ZF precoding in correlated mmWave massive MIMO channels. For a given N_a, ARA are selected to maximize per antenna received power as

$$\text{(P1)} \quad \min_{\mathbf{H}_a \subseteq \mathbf{H}} \quad \text{Tr}\left\{\left(\mathbf{H}_a \mathbf{H}_a^H\right)^{-1}\right\} \tag{6}$$

The detection of spatial and modulation symbols can be recapitulated as follows [6]

- First, the output of the k^{th} AD is compared to $\hat{\gamma}$ to detect k^{th} spatial bit $\hat{s}_{ik} \in \{0, 1\}$ [6].
- Then, the combined signal y_c passes through RF chain to enable detection of the modulation symbol x_j where

$$y_c = \sum_{k=1}^{N_a} \sqrt{\alpha P} \hat{s}_{ik} s_{ik} x_j + \hat{s}_{ik} n_k \tag{7}$$

D. Challenges and proposed solutions

Problem (P1) is solved by exhaustive search in [6] but this method entails considerable computational complexity especially in large MIMO systems. Besides, FD BS shown in Fig. 1 is expensive and power consuming particularly in mmWave massive MIMO systems.

In the sequel, we formulate the joint problem of designing low complexity ZF hybrid precoder and RAS. We divide this problem into two subproblems. First, we select the ARA assuming FD BS by using convex optimization and efficient algorithms to solve (P1). Then, in section VIII we consider those selected antennas in designing novel ZF hybrid precoder. We compare all the proposed designs with the best known.

IV. Joint ZF hybrid precoder design and RAS

At mmWave band, the cost and power consumption of RF chains and high resolution digital-to-analog-converters (DACs) are significant. In this section, we motivate the benefit of the hybrid architecture (Fig. 2) by comparing its power consumption with the FD (Fig. 1) and then we propose a joint problem of ZF hybrid precoder design and RAS.

A. Power consumption

With the goal of justifying the use of hybrid architectures, equation (8) shows the power consumption of power amplifier (P_{PA}), phase shifter (P_{PS}), DAC (P_{DAC}) and RF chain (P_{RF}) in terms of reference power (P_{ref}) [13] as

$$P_{\text{PA}} = P_{\text{ref}}, P_{\text{PS}} = 1.5P_{\text{ref}}, P_{\text{RF}} = 2P_{\text{ref}}, P_{\text{DAC}} = 10P_{\text{ref}} \tag{8}$$

The hardware power consumption of FD BS (P_d) in Fig. 1 and hybrid BS (P_h) in Fig. 2 can be expressed as

$$\begin{aligned} P_h &= N_f N_t \left(P_{\text{PA}} + P_{\text{PS}}\right) + N_f \left(P_{\text{RF}} + P_{\text{DAC}}\right) + P_{\text{BB}} \\ P_d &= N_t \left(P_{\text{PA}} + P_{\text{RF}} + P_{\text{DAC}}\right) + P_{\text{BB}} \end{aligned} \tag{9}$$

where N_f is the number of RF chains and $P_{\text{BB}} = 10P_{\text{ref}}$ is the baseband processing power consumption [13].

Fig. 3 shows hardware power consumption of FD BS and hybrid BS at different values of N_f. At low values of N_f, the hybrid architecture consumes much lower power than FD and the same power as FD at $N_f = 5$. Therefore, hybrid architecture is primness for highly spatially sparse mmWave massive MIMO channels that limited by few scattering clusters.

B. Joint design

The hybrid precoder can be expressed as

$$\mathbf{B}_h = \sqrt{P_g}\mathbf{B}_{\text{RF}}\mathbf{B}_{\text{BB}} = \frac{1}{\| \mathbf{B}_{\text{RF}}\mathbf{B}_{\text{BB}} \|_F}\mathbf{B}_{\text{RF}}\mathbf{B}_{\text{BB}} \tag{10}$$

where $\mathbf{B}_{\text{RF}} \in \mathbb{C}^{N_t \times N_f}$ is the RF precoder that implemented by using phase shifters, $\mathbf{B}_{\text{BB}} \in \mathbb{C}^{N_f \times N_a}$ is the baseband precoder and P_g is the precoder gain.

By considering the equivalent channel $(\mathbf{H}_{\text{eq}} = \mathbf{H}_a\mathbf{B}_{\text{RF}})$, the baseband precoder is designed to zero force the equivalent channel $\left(\mathbf{B}_{\text{BB}} = \mathbf{H}_{\text{eq}}^H \left(\mathbf{H}_{\text{eq}}\mathbf{H}_{\text{eq}}^H\right)^{-1}, N_f \geq N_a\right)$. Hence, the ZF hybrid precoder can be expressed as

$$\mathbf{B}_h = \frac{\mathbf{B}_{\text{RF}}\mathbf{B}_{\text{BB}}}{\sqrt{\text{Tr}\left\{\left(\mathbf{H}_a\mathbf{B}_{\text{RF}}\mathbf{B}_{\text{RF}}^H\mathbf{H}_a^H\right)^{-2}\mathbf{H}_a\left(\mathbf{B}_{\text{RF}}\mathbf{B}_{\text{RF}}^H\right)^2\mathbf{H}_a^H\right\}}} \tag{11}$$

At large N_t, we can assume that $\left(\mathbf{B}_{\text{RF}}^H\mathbf{B}_{\text{RF}} = \mathbf{I}_{N_f}\right)$ and hence, $\mathbf{B}_h$ in equation (11) can be expressed as

$$\mathbf{B}_h = \frac{\mathbf{B}_{\text{RF}}\mathbf{B}_{\text{BB}}}{\sqrt{\text{Tr}\left\{\left(\mathbf{H}_a\mathbf{B}_{\text{RF}}\mathbf{B}_{\text{RF}}^H\mathbf{H}_a^H\right)^{-1}\right\}}} \tag{12}$$

In order to maximize the received signal power, we jointly formulate the RF precoder design and the RAS problems to maximize the precoder gain such as

$$\text{(P2)} \begin{cases} \min\limits_{\mathbf{B}_{\text{RF}}, \mathbf{H}_a \subseteq \mathbf{H}} & \text{Tr}\left\{\left(\mathbf{H}_a\mathbf{B}_{\text{RF}}\mathbf{B}_{\text{RF}}^H\mathbf{H}_a^H\right)^{-1}\right\} \\ \text{s.t.} & \mathbf{B}_{\text{RF}}(n,m) = e^{j\theta_{n,m}},\ \forall n,m. \end{cases} \tag{13}$$

where $j = \sqrt{-1}$. The first step to solve (P2) is to derive a closed form solution to $\mathbf{B}_{\text{RF}}$ assuming $\mathbf{H}_a$ is given. This is difficult because the objective function of (P2) is non-convex, moreover, the constant amplitude of $\mathbf{B}_{\text{RF}}$ is non-convex constraint. We propose to solve (P2) by selecting the ARA at first assuming FD precoder and then we consider those antennas to design the RF precoder.

V. RAS BASED CONVEX OPTIMIZATION

Problem (P1) can be reformulated in terms of eigenvalues of $\left(\mathbf{H}_a\mathbf{H}_a^H\right)$ as

$$\text{(P3)} \begin{cases} \min\limits_{\mathbf{H}_a \subseteq \mathbf{H}} & \sum\limits_{i=1}^{N_a} \frac{1}{\lambda_i} \\ \text{s.t.} & \lambda_i \in \boldsymbol{\lambda}\left\{\mathbf{H}_a\mathbf{H}_a^H\right\}, i = 1, \cdots, N_a. \end{cases} \tag{14}$$

where $\boldsymbol{\lambda}\left\{\mathbf{H}_a\mathbf{H}_a^H\right\}$ is a vector that includes eigenvalues of $\left(\mathbf{H}_a\mathbf{H}_a^H\right)$. Since (P3) is non-convex, we propose two different designs in which we minimize lower bounds on the objective function of (P3). In both cases, we obtain a non-convex problem; however, we relax non-convex constraint to convert into convex program and achieve suboptimal solution.

A. Max-min eigenvalue

The objective function of (P3) is lower bounded by $\frac{1}{\lambda_{N_a}}$ where λ_{N_a} is the smallest eigenvalue of $\left(\mathbf{H}_a\mathbf{H}_a^H\right)$. We propose to minimize this lower bound that implies maximizing λ_{N_a}. The resulting optimization problem can be expressed as

$$\text{(P4)} \begin{cases} \max\limits_{\mathbf{H}_a \subseteq \mathbf{H}} & \lambda_{N_a} \\ \text{s.t.} & \lambda_i \in \boldsymbol{\lambda}\left\{\mathbf{H}_a\mathbf{H}_a^H\right\}, \\ & \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{N_a}. \end{cases} \tag{15}$$

Let us define $\mathbf{X} \in \mathbb{R}^{N_r \times N_r}$ as a diagonal matrix and $\mathbf{X}_i$ is the i^{th} diagonal element that follows

$$\mathbf{X}_i = \begin{cases} 1 & \text{if } i^{\text{th}} \text{ receive antenna is active} \\ 0 & \text{if } i^{\text{th}} \text{ receive antenna is silent} \end{cases} \tag{16}$$

Let us assume that matrices $\mathbf{X}$ and $\mathbf{H}_a$ share the same N_a ARA. Thus, the largest N_a eigenvalues of $\left(\mathbf{H}^H\mathbf{X}\mathbf{H}\right)$ are same as the eigenvalues of $\mathbf{H}_a\mathbf{H}_a^H$. Moreover, the smallest $(N_t - N_a)$ eigenvalues of $\left(\mathbf{H}^H\mathbf{X}\mathbf{H}\right)$ are zeros. Therefore, without loss of optimality, problem (P4) can be expressed as

$$\text{(P5)} \begin{cases} \max\limits_{\mathbf{X}} & \sum\limits_{i=N_a}^{N_t} \lambda_i \\ \text{s.t.} & \lambda_i \in \boldsymbol{\lambda}\left\{\mathbf{H}^H\mathbf{X}\mathbf{H}\right\}, \\ & \lambda_1 \geq \lambda_2 \geq \cdots > \lambda_{N_t}. \\ & \mathbf{X} \in \text{diagonal},\ \mathbf{X}_i \in \{0,1\}. \\ & \text{Tr}\{\mathbf{X}\} = N_a. \end{cases} \tag{17}$$

Although the objective function of (P5) is concave in $\mathbf{X}$, (P5) is non-convex optimization problem because $(\mathbf{X}_i \in \{0,1\})$ is non-convex constraint. We do relaxation to convert (P5) into convex optimization problem where replace the non-convex constraint with linear one $(0 \leq \mathbf{X}_i \leq 1)$.

Solution $\mathbf{X}^\star$ of the relaxed problem does not follow equation (16). Therefore, we generate another solution $\underline{\mathbf{X}}^\star$ such that has ones in positions of maximum N_a diagonal elements of $\mathbf{X}^\star$ and zeros in the other locations.

B. Min sum of convex fractions

We propose an alternative formulation of the problem using a tighter bound. The proposed lower bound satisfy the following inequality

$$\frac{1}{\lambda_{N_a}} \leq \sum_{i=1}^{N_a} \frac{1}{\sum_{j=N_a-i+1}^{N_a} \lambda_j} \leq \sum_{i=1}^{N_a} \frac{1}{\lambda_i} \tag{18}$$

where $\lambda_i \in \boldsymbol{\lambda}\left\{\mathbf{H}_a\mathbf{H}_a^H\right\}, \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{N_a}$. Consequently, the problem can be expressed as

$$\text{(P6)} \begin{cases} \min\limits_{\mathbf{H}_a \subseteq \mathbf{H}} & \sum\limits_{i=1}^{N_a} \frac{1}{\sum_{j=N_a-i+1}^{N_a} \lambda_j} \\ \text{s.t.} & \lambda_i \in \boldsymbol{\lambda}\left\{\mathbf{H}_a\mathbf{H}_a^H\right\}, \\ & \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{N_a}. \end{cases} \tag{19}$$

The fractional nature of (P6) can be exploited to solve this problem. In optimization theory, fractional programming (FP) [14] provides low computational complexity algorithms such as Dinkelbach algorithm [15] to minimize sum of fractional functions of convex numerator and concave denominator (convex fraction). Problem (P6) can be expressed as sum of convex fractions

$$\text{(P7)} \begin{cases} \min\limits_{\mathbf{X}} & \sum\limits_{i=1}^{N_a} \frac{1}{\sum_{j=N_a-i+1}^{N_a} \lambda_j} \\ \text{s.t.} & \lambda_i \in \boldsymbol{\lambda}\left\{\mathbf{H}^H\mathbf{X}\mathbf{H}\right\}, \\ & \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{N_t}. \\ & \mathbf{X} \in \text{diagonal},\ 0 \leq \mathbf{X}_i \leq 1. \\ & \text{Tr}\{\mathbf{X}\} = N_a. \end{cases} \tag{20}$$

where each fraction in the objective function of (P7) is convex fraction that has constant numerator and concave denominator. In Algorithm 1, we convert the non-convex (P7) into sequence of convex problems. Moreover, Algorithm 1 converges in few iterations.

Algorithm 1 RAS via Dinkelbach algorithm [15]

Input : $\epsilon < 0$, $\rho_i = 1$, for all $i = 1, \cdots N_a$.
Output : $\underset{\sim}{\mathbf{X}}^\star$
repeat

$$\underset{\sim}{\mathbf{X}}^\star = \begin{cases} \min\limits_{\mathbf{X}} & N_a - \sum\limits_{i=1}^{N_a} \rho_i \sum\limits_{j=N_a-i+1}^{N_t} \lambda_j \\ \text{s.t.} & \lambda_i \in \boldsymbol{\lambda}\left\{\mathbf{H}^H\mathbf{X}\mathbf{H}\right\}, \\ & \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{N_t}. \\ & \mathbf{X} \in \text{diagonal},\ 0 \leq \mathbf{X}_i \leq 1. \\ & \text{Tr}\left\{\mathbf{X}\right\} = N_a. \end{cases}$$

$F\left(\underset{\sim}{\mathbf{X}}^\star\right) = N_a - \sum_{i=1}^{N_a} \rho_i \sum_{j=N_a-i+1}^{N_t} \lambda_i$
$\rho_i = \frac{1}{\sum_{j=N_a-i+1}^{N_t} \lambda_i}, \quad i = 1, \cdots, N_a.$
until $F\left(\underset{\sim}{\mathbf{X}}^\star\right) > \epsilon$

VI. Sequential add/remove algorithms

The introduced lower bounds in (P5)-(P7) and the relaxation lead to suboptimal solutions. In this section, we propose RAS designs that approach the optimal selection by using efficient sequential methods to solve (P3). In Algorithms 2 and 3, we (activate/deactivate) one antenna per iteration that has major (positive/negative) effect on (P3). In Algorithms 4 and 5, we reduce the complexity of Algorithms 2 and 3 by replacing eigenvalues computation with simple projection methods. Finally, we compare all proposed algorithms with optimal selection (exhaustive search) in terms of performance and complexity.

A. Best empty initialization

In Algorithm 2, we propose efficient iterative method to solve (P3) where we select one good receive antenna per iteration. Initially, $(\mathbf{X} = \mathbf{0}_{N_r})$ and during the i^{th} iteration the specific diagonal entry in $\mathbf{X}$ is set to one to minimize $\left(\frac{1}{\lambda_1} + \cdots + \frac{1}{\lambda_i}\right)$ where $(\lambda_1, \cdots, \lambda_i)$ are the largest i eigenvalues of $\mathbf{H}^H\mathbf{X}\mathbf{H}$.

B. Best full initialization

Algorithm 3 shows efficient sequential technique to solve (P3) where we deactivate one bad receive antenna per iteration. Initially, $(\mathbf{X} = \mathbf{I}_{N_r})$ and during the i^{th} iteration specific diagonal entry in $\mathbf{X}$ be equal to zero to minimize $\left(\frac{1}{\lambda_1} + \cdots + \frac{1}{\lambda_{N_a}}\right)$ where $(\lambda_1, \cdots, \lambda_{N_a})$ are the largest N_a eigenvalues of $\mathbf{H}^H\mathbf{X}\mathbf{H}$.

With the aim of avoiding the complexity resulting from eigenvalues computation in Algorithms 2 and 3, Algorithms 4 and 5 show iterative designs to select the ARA based simple projection methods.

C. Low complexity empty initialization

In Algorithm 4, $\mathbf{X} = \mathbf{0}_{N_r}$ at initial. Then, we select the receive antenna that has channel vector with the highest norm. Next, at the i^{th} iteration we select one receive antenna where its channel vector $\mathbf{H}\left(i,:\right)^H$ has minimum projection on $\mathbf{H}^H\mathbf{X}\mathbf{H}$.

Algorithm 2 Best RAS via empty initialization

Input : $(\mathbf{H}, N_r, N_a)$
Output : $\mathbf{H}_a$
$\mathbf{X} = \mathbf{0}_{N_r}, \mathbf{Y} = \mathbf{0}_{N_r}$
$\mathcal{V} = \{1, \cdots, N_r\}, \mathcal{J} = \{1, \cdots, N_r\}, \mathcal{K} =$ Empty set
for $i = 1 : N_a$
  for $j \in \mathcal{V}$
    $\mathbf{X}\left(j,j\right) = 1$
    $\lambda_1, \cdots, \lambda_i$ are largest i eigenvalues of $\mathbf{H}^H\mathbf{X}\mathbf{H}$
    $l_j = \sum_{k=1}^{i} \frac{1}{\lambda_k}$
    $\mathbf{X} = \mathbf{Y}$
  end for
  $\mathcal{K} = \mathcal{K} + \arg\min\limits_{\forall j} \quad l_j$
  $\mathbf{X}\left(\mathcal{K}\right) = 1, \mathbf{Y}\left(\mathcal{K}\right) = 1, \mathcal{V} = \mathcal{J} - \mathcal{K}$
end for
return $\mathbf{H}_a = \mathbf{H}(\mathcal{K}, :)$

Algorithm 3 Best RAS via full initialization

Input : $(\mathbf{H}, N_r, N_a)$
Output : $\mathbf{H}_a$
$\mathbf{X} = \mathbf{I}_{N_r}, \mathbf{Y} = \mathbf{I}_{N_r}$
$\mathcal{V} = \{1, \cdots, N_r\}, \mathcal{J} = \{1, \cdots, N_r\}, \mathcal{K} =$ Empty set
for $i = 1 : N_r - N_a$
  for $j \in \mathcal{V}$
    $\mathbf{X}\left(j,j\right) = 0$
    $\lambda_1, \cdots, \lambda_{N_a}$ are largest N_a eigenvalues of $\mathbf{H}^H\mathbf{X}\mathbf{H}$
    $l_j = \sum_{k=1}^{N_a} \frac{1}{\lambda_k}$
    $\mathbf{X} = \mathbf{Y}$
  end for
  $\mathcal{K} = \mathcal{K} + \arg\min\limits_{\forall j} \quad l_j$
  $\mathbf{X}\left(\mathcal{K}\right) = 0, \mathbf{Y}\left(\mathcal{K}\right) = 0, \mathcal{V} = \mathcal{J} - \mathcal{K}$
end for
return $\mathbf{H}_a = \mathbf{H}(\mathcal{V}, :)$

D. Low complexity full initialization

In Algorithm 5, we start with $\mathbf{X} = \mathbf{I}_{N_r}$. Then, at the i^{th} iteration we deactivate one receive antenna where its channel vector has maximum projection on $\mathbf{H}^H\mathbf{X}\mathbf{H}$.

In Fig. 4, we compare all proposed RAS designs with optimal selection in terms of objective function of (P1). RAS designs based convex optimization have lower performance than those based sequential algorithms because the relaxation of the non-convex constraint $(\mathbf{X}_i \in \{0, 1\})$. The lower bound in (P6) is tighter than the lower bound in (P4). Therefore, Algorithm 1 outperforms the design in (P5). Algorithms 2 and 3 approach optimal selection with much lower computational complexity. Algorithms 4 and 5 avoid the complexity resulting from eigenvalues computation in exchange for less performance than Algorithms 2 and 3. Maximum and minimum initialization are approaching in performance. However, each is computationally efficient at certain range of N_a.

Algorithm 4 Low complexity RAS via empty initialization

Input : $(\mathbf{H}, N_r, N_a)$
Output : $\mathbf{H}_a$
$\mathbf{X} = \mathbf{0}_{N_r}, \mathcal{J} = \{1, \cdots, N_r\}, \mathcal{K} = \text{Empty set}$
$k = \arg\max_{\forall j} \text{diag}\left(\mathbf{H}\mathbf{H}^H\right)$
$\mathbf{X}(k,k) = 1, \mathcal{K} = \mathcal{K} + k, \mathcal{V} = \mathcal{J} - \mathcal{K}$
for $i = 2 : N_a$
  for $j \in \mathcal{V}$
    $\mathbf{z} = \mathbf{H}(j,:)^H$
    $l_j = \|\mathbf{H}^H\mathbf{X}\mathbf{H}\mathbf{z}\|_2$
  end for
  $\mathcal{K} = \mathcal{K} + \arg\min_{\forall j} \quad l_j$
  $\mathbf{X}(\mathcal{K}) = 1, \mathcal{V} = \mathcal{J} - \mathcal{K}$
end for
return $\mathbf{H}_a = \mathbf{H}(\mathcal{K},:)$

Algorithm 5 Low complexity RAS via full initialization

Input : $(\mathbf{H}, N_r, N_a)$
Output : $\mathbf{H}_a$
$\mathbf{X} = \mathbf{I}_{N_r}, \mathbf{Y} = \mathbf{I}_{N_r}$
$\mathcal{V} = \{1, \cdots, N_r\}, \mathcal{J} = \{1, \cdots, N_r\}, \mathcal{K} = \text{Empty set}$
for $i = 1 : N_r - N_a$
  for $j \in \mathcal{V}$
    $\mathbf{X}(j,j) = 0$
    $\mathbf{z} = \mathbf{H}(j,:)^H$
    $l_j = \|\mathbf{H}^H\mathbf{X}\mathbf{H}\mathbf{z}\|_2$
  end for
  $\mathcal{K} = \mathcal{K} + \arg\max_{\forall j} \quad l_j$
  $\mathbf{X}(\mathcal{K}) = 0, \mathbf{Y}(\mathcal{K}) = 0, \mathcal{V} = \mathcal{J} - \mathcal{K}$
end for
return $\mathbf{H}_a = \mathbf{H}(\mathcal{V},:)$

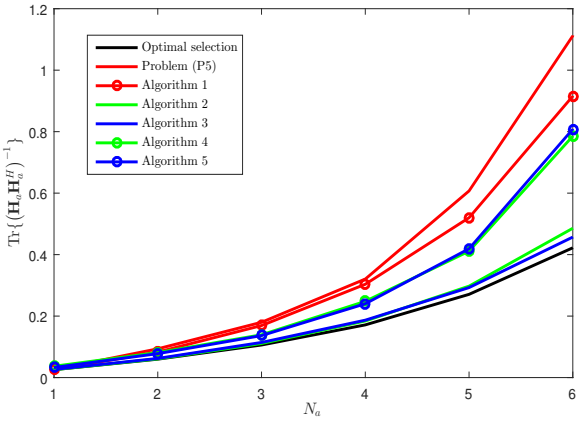

Fig. 4. Objective function of (P1) versus N_a at 16×32 mmWave channel in (1), $\xi = 1$, $L = 16$ and (100 channel realizations).

E. Computational complexity analysis

The number of singular-value-decomposition (SVD) needed for optimal selection (S_o), empty initialization of Algorithm 2 (S_e) and full initialization of Algorithm 3 (S_f) can be expressed as

$$\begin{aligned} S_o &= \binom{N_r}{N_a} \\ S_e &= \sum_{i=0}^{N_a-1} N_r - i = N_a\left(N_r - \frac{1}{2}(N_a - 1)\right) \\ S_f &= \sum_{i=0}^{N_r-N_a-1} N_r - i \\ &= (N_r - N_a)\left(N_r - \frac{1}{2}(N_r - N_a - 1)\right) \end{aligned} \tag{21}$$

In terms of computational complexity, equation (21) shows that empty initialization is the best when $N_a < \frac{N_r}{2}$ while full initialization is better when $N_a > \frac{N_r}{2}$.

As an illustrative example, Table I shows numerical values of S_o, S_e and S_f at $N_r = 16$ and different numbers of ARA. Therefore, the proposed algorithms are computationally simpler than optimal selection and closely approach the optimal performance.

TABLE I
SVD OPERATIONS NEEDED AT $N_r = 16$

N_a	S_o	S_e	S_f
6	8008	81	115
8	12870	100	100
10	8008	115	81

VII. MUTUAL INFORMATION AND OPTIMAL N_a

In this section, we derive the mutual information of the discrete channel implemented in the RSM system in Fig. 1. Then, we show that the mutual information as a function of N_a has one global maximum. Subsequently, we propose fast algorithm to find the optimal value of N_a that maximizes the mutual information.

According to equations (3)-(5)-(7), the mutual information between the transmitted and the received symbols can be expressed by applying the chain rule in [16] as

$$\begin{aligned} I(\mathbf{s}, x; \hat{\mathbf{s}}, y) &= I(\mathbf{s}, x; \hat{\mathbf{s}}) + I(\mathbf{s}, x; y|\hat{\mathbf{s}}) \\ I(\mathbf{s}, x; \hat{\mathbf{s}}) &= I(\mathbf{s}; \hat{\mathbf{s}}) + I(x; \hat{\mathbf{s}}|\mathbf{s}) \\ I(\mathbf{s}, x; y|\hat{\mathbf{s}}) &= I(\mathbf{s}; y|\hat{\mathbf{s}}) + I(x; y|\hat{\mathbf{s}}, \mathbf{s}) \end{aligned} \tag{22}$$

Since $\hat{\mathbf{s}}$ is used only for spatial symbol detection and y for modulation symbol detection, $(I(x; \hat{\mathbf{s}}|\mathbf{s}) = 0, I(\mathbf{s}; y|\hat{\mathbf{s}}) = 0)$. Hence, the mutual information can be expressed as

$$I(\mathbf{s}, x; \hat{\mathbf{s}}, y) = I(\mathbf{s}; \hat{\mathbf{s}}) + I(x; y|\hat{\mathbf{s}}, \mathbf{s}) \triangleq I_s + I_m \tag{23}$$

According to equation (5), the AD connected to each ARA can receive one of two amplitudes $|n|$ or $|\sqrt{\alpha P}+n|$. Then, the measured amplitude is compared with $\hat{\gamma}$ to detect one spatial bit per antenna. Therefore, spatial mutual information (I_s) can

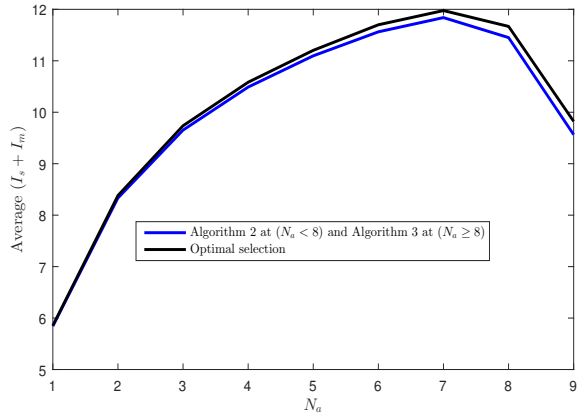

Fig. 5. Average mutual information versus N_a at 16×32 mmWave channel in (1), $\xi = 1$, $L = 16$, $\frac{P}{\sigma^2} = 10$dB and (100 channel realizations).

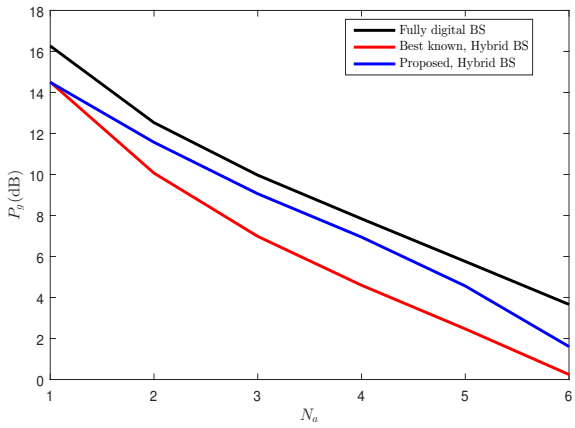

Fig. 6. Zero forcing precoder gain versus N_a at 16×32 mmWave channel in (1), $\xi = 1$, $L = 16$, $N_f = N_a + 1$ and (100 channel realizations).

be computed using the binary asymmetric channel [17] as

$$I_s = N_a \left(\mathcal{H}\left(\frac{P_1 + 1 - P_0}{2} \right) - \frac{\mathcal{H}(P_1) + \mathcal{H}(1 - P_0)}{2} \right)$$
$$P_1 = \Pr\left(|\sqrt{\alpha P} + n| > \hat{\gamma} \right) = Q_1\left(\frac{1}{\sigma}\sqrt{2\alpha P}, \frac{1}{\sigma}\sqrt{2}\hat{\gamma} \right)$$
$$P_0 = \Pr(|n| < \hat{\gamma}) = 1 - Q_1\left(0, \frac{1}{\sigma}\sqrt{2}\hat{\gamma} \right), \hat{\gamma} \approx \frac{1}{2}\sqrt{\alpha P} \quad (24)$$

where $\mathcal{H}(P) = -P\log_2 P - (1-P)\log_2(1-P)$ and $Q_1(x)$ is first order marcum Q function [18].

Since the BS also transmits one modulation symbol(assume Gaussian), the modulation mutual information (I_m) can be characterized by multiple-input-single-output channel as

$$I_m = \sum_{i=1}^{2^{N_a}} \Pr(\mathbf{s}_i) \sum_{j=1}^{2^{N_a}} \Pr(\hat{\mathbf{s}}_j | \mathbf{s}_i) \log_2\left(1 + \mathrm{SNR}|_{\mathbf{s}_i, \hat{\mathbf{s}}_j}\right) \quad (25)$$

$$\mathrm{SNR}|_{\mathbf{s}_i, \hat{\mathbf{s}}_j} = \frac{\left(\sum_{k=1}^{N_a} s_{ik} \hat{s}_{jk} \right)^2}{\max\left(\sum_{k=1}^{N_a} \hat{s}_{jk}, 1 \right)} \frac{\alpha P}{\sigma^2} \quad (26)$$

$$\Pr(\hat{\mathbf{s}}_j | \mathbf{s}_i) = \prod_{k=1}^{N_a} \Pr\left(|y_{ik}| \underset{\hat{s}_{jk}=0}{\overset{\hat{s}_{jk}=1}{\gtrless}} \hat{\gamma} \right), \Pr(\mathbf{s}_i) = \frac{1}{2^{N_a}} \quad (27)$$

Fig. 5 shows that average mutual information achieved by proposed fast algorithms approach the one obtained by optimal selection. The mutual information increases with N_a until it reaches the maximum then it decreases. Therefore, we can find the optimal N_a by fast iterative algorithm that starts with $N_a = 1$ and stops when mutual information decreases.

VIII. ZF Hybrid Precoder

After selecting the ARA, we propose novel ZF RF precoder design and we prove that the ZF hybrid precoder is the same as ZF FD precoder at channels with high spatial sparsity. We show that the proposed precoder outperforms the best known in the literature in performance and complexity.

We solve (P2) assuming $\mathbf{H}_a$ is known. We drop the constant amplitude constraint of problem (P2) and replace the quadratic term $\mathbf{B}_{\mathrm{RF}}\mathbf{B}_{\mathrm{RF}}^H$ by the linear term $\mathbf{Y}$ to relax (P2) into a convex problem that can be expressed as

$$(\mathrm{P9}) \begin{cases} \min\limits_{\mathbf{Y}} & \mathrm{Tr}\left\{ \left(\mathbf{H}_a \mathbf{Y} \mathbf{H}_a^H \right)^{-1} \right\} \\ \text{s.t.} & \mathrm{Tr}\{\mathbf{Y}\} = N_t N_f. \\ & \mathbf{Y} \succeq 0 \end{cases} \quad (28)$$

Solution $\mathbf{Y}$ has arbitrary rank profile so the RF precoder is designed based the largest N_f eigenvectors of $\mathbf{Y}$ as

$$\mathbf{B}_{\mathrm{RF}} = \mathrm{Arg}\left(\mathbf{V}_{N_f}\{\mathbf{Y}\} \right) \quad (29)$$

A. High spatial sparsity $(L \leq N_f)$

The channel matrix in equation (1) can be expressed as

$$\mathbf{H} = \mathbf{A}_r \mathbf{D} \mathbf{A}_t^H \quad (30)$$

where $\mathbf{D} \in \mathbb{C}^{L \times L}$ is the path gain diagonal matrix, $\mathbf{A}_r \in \mathbb{C}^{N_r \times L}$ and $\mathbf{A}_t \in \mathbb{C}^{N_t \times L}$ are matrices containing receive and transmit response vectors, respectively. After RAS, the channel matrix $\mathbf{H}_a$ can be expressed as

$$\mathbf{H}_a = \mathbf{A}_r(\mathcal{S}, :) \mathbf{D} \mathbf{A}_t^H = \mathbf{A}_{ra} \mathbf{D} \mathbf{A}_t^H \quad (31)$$

where $\mathcal{S}$ is set contains indices of ARA.

If $L \leq N_f$, ZF hybrid precoder $\mathbf{B}_h$ becomes exactly the same as ZF digital precoder $\mathbf{B}_d$ because there is a unique RF chain for each scattering path as illustrated in equation (32).

$$\begin{aligned} \mathbf{B}_d &= \mathbf{H}_a^H \left(\mathbf{H}_a \mathbf{H}_a^H \right)^{-1} = (\mathbf{A}_t) \left(\mathbf{D} \mathbf{A}_{ra}^H \left(\mathbf{H}_a \mathbf{H}_a^H \right)^{-1} \right) \\ &= (\mathbf{B}_{\mathrm{RF}})(\mathbf{B}_{\mathrm{BB}}) = \mathbf{B}_h, \quad N_a \leq L \end{aligned} \quad (32)$$

B. Best known ZF RF precoder

In Algorithm 3 in [10], the authors proposed iterative method to solve (P2) assuming $\mathbf{H}_a$ is known. In this algorithm; at first, $\mathbf{B}_{\mathrm{RF}}$ is any feasible matrix. Then, one element of $\mathbf{B}_{\mathrm{RF}}$ is updated per iteration. The algorithm stops when $\mathbf{B}_{\mathrm{RF}}$ converges. However, this design is computationally complex

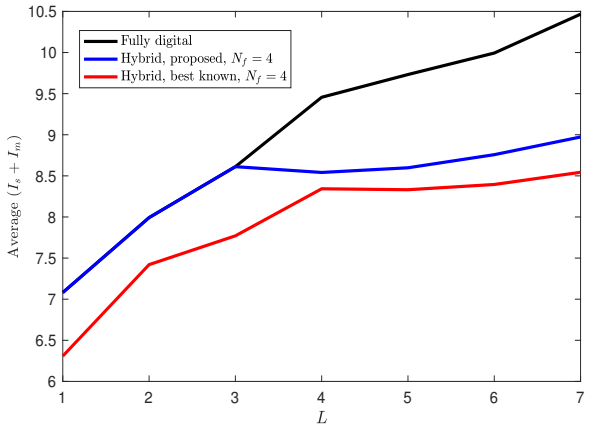

Fig. 7. Average mutual information versus L at 16×32 mmWave channel in (1), $\xi = 1$, $\frac{P}{\sigma^2} = 10$dB and (100 channel realizations).

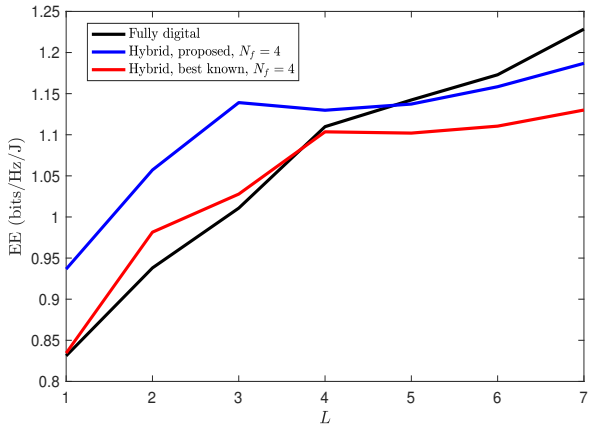

Fig. 8. Energy efficiency versus L at 16×32 mmWave channel in (1), $\xi = 1$, $P_{\text{ref}} = 20$mW, $\frac{P}{\sigma^2} = 10$dB and (100 channel realizations).

because it needs at least $N_t N_f$ iterations to reconstruct $\mathbf{B}_{\text{RF}}$. Moreover, it needs $N_f \geq N_a + 1$; on contrary, the proposed design in (29) needs $N_f \geq N_a$. The proposed precoder not only simpler than the design in [10] but also it achieves higher precoding gain as illustrated in Fig. 6.

Fig. 7 shows the average mutual information of RSM system considering several precoders. For all precoding schemes, N_a is selected to maximize the mutual information. The proposed hybrid precoder is not only superior to that based the design in [10] but also equal to that based FD precoder when $(L \leq N_f)$. The standard deviations of the mutual information over the 100 realizations are close to one for all precoders.

In Fig. 8, energy efficiency (EE) is defined as mutual information per BS power consumption. The proposed hybrid precoding scheme is the most energy efficient architecture when the channel is highly spatially sparse. On the other hand, FD precoding architecture becomes more energy efficient when sparsity level decreases.

IX. Conclusion

The proposed RAS based convex optimization are suboptimal due to relaxing non-convex constraints. Therefore, performance of proposed RAS sequential algorithms are superior to that based convex optimization. For low computational complexity, its useful to start with empty initialization when $\left(N_a < \frac{N_r}{2}\right)$ and full initialization at $\left(N_a \geq \frac{N_r}{2}\right)$. We also developed fast algorithm to determine the optimal number of ARA that maximizes the mutual information. The proposed ZF hybrid precoder outperforms the best known design and becomes optimal when the channel is very spatially sparse. Hybrid precoder is the most energy efficient when the channel is limited by few number of scattering paths; otherwise, FD is better. Optimizing (P2) and designing low complexity sub-connected hybrid precoders are future work topics.

References

[1] R. W. Heath et al., "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, Feb. 2016.

[2] Lie-Liang Yang, "Transmitter preprocessing aided spatial modulation for multiple-input multiple-output systems," *in 73rd IEEE Vehicular Technology Conference (VTC Spring)*, pp. 1–5, May 2011.

[3] A. Stavridis, S. Sinanovic, M. Di Renzo, and H. Haas, "Transmit precoding for receive spatial modulation using imperfect channel knowledge," *in 75th IEEE Vehicular Technology Conference (VTC Spring)*, pp. 1–5, May 2012.

[4] R. Zhang, Lie-Liang Yang, and L. Hanzo, "Generalised pre-coding aided spatial modulation," *IEEE Transactions on Wireless Communications*, vol. 12, no. 11, pp. 5434–5443, Nov. 2013.

[5] N. S. Perovic, P. Liu, M. Di Renzo, and A. Springer, "Receive spatial modulation for LOS mmwave communications based on TX beamforming," *IEEE Communications Letters*, Dec. 2016.

[6] A. Raafat, A. Agustin, and J. Vidal, "Receive spatial modulation for massive MIMO systems," *in IEEE Global Communications Conference (GLOBECOM)*, Dec. 2017.

[7] M. Gharavi-Alkhansari and A. B. Gershman, "Fast antenna subset selection in MIMO systems," *IEEE transactions on signal processing*, vol. 52, no. 2, pp. 339–347, Feb. 2004.

[8] A. Dua, K. Medepalli, and A. J. Paulraj, "Receive antenna selection in MIMO systems using convex optimization," *IEEE Transactions on Wireless Communications*, vol. 5, no. 9, pp. 2353–2357, Sept. 2006.

[9] 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, Dec. 2014.

[10] 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, April 2016.

[11] M. R. Akdeniz et al., "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE journal on selected areas in communications*, vol. 32, no. 6, pp. 1164–1179, June 2014.

[12] S. Rami, W. Tuni, and W. R. Eisenstadt, "Millimeter wave MOSFET amplitude detector," *Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems (SIRF)*, pp. 84–87, Jan. 2010.

[13] R. Méndez-Rial et al., "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?," *IEEE Access*, vol. 4, pp. 247–267, Jan. 2016.

[14] S. Schaible and T. Ibaraki, "Fractional programming," *European Journal of Operational Research*, vol. 12, no. 4, pp. 325–338, April 1983.

[15] Y. Almogy and O. Levin, "A class of fractional programming problems," *Operations Research*, vol. 19, no. 1, pp. 57–67, Feb. 1971.

[16] T. M. Cover and J. A. Thomas, *Elements of information theory*, John Wiley & Sons, 2012.

[17] F. Chapeau-Blondeau, "Noise-enhanced capacity via stochastic resonance in an asymmetric binary channel," *Physical Review E*, vol. 55, no. 2, pp. 2016, Feb. 1997.

[18] M. K. Simon, *Probability distributions involving Gaussian random variables: A handbook for engineers and scientists*, Springer, 2007.