

Energy Efficient Transmit-Receive Hybrid Spatial Modulation for Large-Scale MIMO Systems

Ahmed Raafat, Merve Yüzgeçcioğlu, *Student Member, IEEE*, Adrian Agustin, *Member, IEEE*, Josep Vidal, *Senior Member, IEEE*, Eduard A. Jorswieck, *Senior Member, IEEE* and Yoann Corre, *Member, IEEE*

Abstract—We consider a point to point large-scale multiple-input multiple-output system operating in the millimeter wave (mmWave) band and an outdoor scenario. Novel transmit and receive spatial modulation schemes are proposed for uplink (UL) and downlink (DL) data transmission phases based on a novel energy efficient hybrid user terminal architecture. The analog circuitry of the proposed hybrid architecture is divided into two stages: phase shifters and analog switches. The phase shifting stage assures high gain and overcomes the severe path-loss caused by outdoor mmWave propagation. The analog switching stage smartly allocates the antennas to be used at the phase shifting stage and combats the spatial correlation. We provide the analysis of the spectral efficiency (SE) of the UL and DL systems. Next, we propose a reduced complexity algorithm that jointly optimizes the analog beamformer and combiner design of the UL and DL circuitry to maximize the energy efficiency (EE). Finally, we compare and evaluate the performance of the proposed algorithm in terms of the SE and EE assuming both stochastic and realistic channel models.

I. INTRODUCTION

WITH the advances in communications technology and the new wireless-based applications and services, it is forecasted that the number of connected devices will reach tens of billions by 2030 [1]. In addition, every device connected to the network will demand an average data rate increase of 1000x compared to today's networks. The shortage of bandwidth in low frequency bands has led to the consideration of migrating to the higher frequencies (mmWave) bands, thus attracting the research community as a way to deliver expected large traffic demands [2], [3].

Shortcomings of the mmWave band are the severe path-loss and sensitivity to blockage [4]. Large-scale multiple-input multiple output (MIMO) technology is one of the promising candidate technologies to combat these challenges [5]. By increasing the number of antennas, better performance can be

achieved in terms of a trade-off between antenna gain and degrees of freedom [6]. Linear precoding schemes can be employed to reduce the complexity of the system drastically [7]. Having accurate knowledge of the channel state information (CSI) in transmission is crucial in the design of the precoder that nulls out the unintended space-time-frequency dimensions. A larger number of antennas comes together with an increased complexity and bandwidth usage required to precisely acquire CSI. A pilot-based channel estimation using polynomial expansion proposed in [8] and a low complexity adaptive compress sensing based algorithm proposed in [9] are two among the many low complexity methods proposed to estimate the CSI of large-scale MIMO systems.

Fully digital architectures such as block diagonalization algorithm in [10] have been proven to achieve high performance. However, implementation of a fully digital architecture at a terminal with large number of antennas is not practical due to space limitations and energy consumption constraints. Therefore, hybrid architectures have raised interest as an attractive solution that allows designs with a reduced number of radio frequency (RF) chains by combining analog beamforming together with digital precoding [11], [12]. Various algorithms have been proposed that configure fully connected hybrid precoders at the transmitter side and analog combiner at the receiver end with a small training and feedback overhead [13]. In this study, the authors have shown that the hybrid beamforming system achieves higher data rate values compared to the analog beamforming and closely approach the performance of fully digital beamforming. However, the fully connected hybrid MIMO architecture proposed there comprises many analog devices (phase shifters, power splitters and combiners) which entail higher power consumption than fully digital MIMO designs [14], [15]. Abandoning the fully connected hybrid architecture results in degradation of the spectral efficiency (SE) even with the use of the most sophisticated algorithms in the design the sub-connected hybrid precoders and combiners as illustrated in Fig. 10 in [14] and Fig. 3 in [16]. In [14], the authors studied the SE – EE trade-off of sub-connected hybrid architectures that consist of phase shifters or switches. They showed that a better EE is attained with the proposed hybrid architecture at the cost of losing SE when compared with the fully connected architectures.

Spatial modulation (SM) techniques have been initially introduced for sub-6 GHz with the aim of reducing the number of RF chains and minimizing the hardware cost and power consumption [17]. They are able to enhance SE by exploiting the spatial dimension. The concept can be applied

A. Raafat, A. Agustin and J. Vidal are with the Department of Signal Theory and Communications, Universitat Politècnica de Catalunya, Barcelona 08034, Spain (e-mail: ahmed.raafat@upc.edu; adrian.agustin@upc.edu; josep.vidal@upc.edu).

M. Yüzgeçcioğlu and E. A. Jorswieck are with the Department of Electrical Engineering and Information Technology, TU Dresden, 01062 Dresden, Germany (e-mail: merve.yuzgeccioğlu@tu-dresden.de ; eduard.jorswieck@tu-dresden.de).

Y. Corre is with Siradel, 35043 Rennes Cedex, France (e-mail: ycorre@siradel.com).

The research leading to these results has been funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 641985, the project 5G&B RUNNER-UPC (TEC2016-77148-C2-1-R (AEI/FEDER, UE)) and the Catalan Government (2017 SGR 578 - AGAUR).

at the transmitter (transmit SM (TSM)) [17] or at the receive (receive SM (RSM)) [18]. In TSM, one antenna is active during the transmission and the remaining are silent and thus, only one RF chain is needed. Part of the input data bits is mapped into the index of the active transmit antenna and the other part is mapped into M -ary modulation symbol to be transmitted from the active antenna. After that, the receiver applies maximum likelihood (ML) detector to jointly detect the index of the active transmit antenna and the M -ary symbol assuming that the CSI is available at the receiver. For the sake of improving the SE, generalized SM (GSM) techniques have been developed to enable activating a set of transmit antennas instead of one transmit antenna for the TSM systems [19]. However, TSM schemes suffer from low antenna gain because most of the transmit antennas are silent which is an impairment at higher frequencies (28 GHz) where large beamforming gain is needed to combat the severe path-loss. Hybrid TSM (HTSM) schemes have been reported in [20], [21] to exploit the phase shifters in attaining high beamforming gain where SM bits are mapped into a group of antennas each connected to phase shifter instead of single antenna of TSM. However, outdoor propagation of mmWave signals exhibits poor scattering which entails spatially sparse and rank deficient channel matrices as explained in Sec. V-C and Fig. 10 in [22]. Thus, the performance of HTSM schemes in [20], [21] can be highly degraded if the receiver cannot distinguish between correlated phase shifters groups. On the other hand, in RSM, a subset of receive antennas are active during a transmission. In contrast to the previous case, the SM bits are devoted to indicate the active receive antennas whilst the other bits are mapped into M -ary symbol [18]. In RSM, the transmitter applies zero forcing (ZF) precoding assuming CSI knowledge and the receiver applies the ML principle to jointly detect the index of the active receive antennas set and the M -ary symbol. In the literature, RSM systems adopt fully digital MIMO architectures and thus exhibit high power consumption especially for systems operating at mmWave band. Moreover, ZF precoders suffer from performance degradation in rank deficient MIMO channels. In [23], the authors developed an RSM scheme for indoor propagation of mmWave signals where they control the inter antenna spacing to ensure orthogonal MIMO channel. However, the proposal is not practical in outdoor scenarios as the large transmitter-receiver separation requires large inter antenna spacing. In [24], the authors developed HTSM in the uplink (UL) and RSM in the downlink (DL) assuming outdoor propagation of mmWave signals. However, they do not consider a hybrid structure for the RSM in DL and rely on a rigid hybrid architecture for the UL TSM. In [25], [26], the authors developed HTSM based on partially connected hybrid transmitter and fully digital receiver. The works in [24]-[25], [26], consider that the phase shifting groups, the number of phase shifters per group and the set of antennas per the phase shifting group are fixed which is clearly suboptimal in terms of EE and SE.

In this paper, we consider a point to point¹ large-scale

MIMO system that operates in a mmWave outdoor narrowband channel scenario and we tackle the drawbacks of the SM techniques (small antenna gain and MIMO channel rank deficiency). The fully connected hybrid MIMO depends on large number of phase shifters. The power consumption of the switch is much less than that of the phase shifter. Thus, we consider user terminal (UT) architecture consists of partially connected phase shifters network and switches network. Specifically, we propose an energy efficient UL and DL hybrid design that adopts TSM during the UL and RSM during the DL. In order to cope with the small antennas gain of SM schemes, we consider a hybrid architecture in both UL and DL to maintain high beamforming and combining gains, respectively and combat the severe path-loss of the outdoor mmWave propagation. In order to ensure full rank MIMO channel condition, we propose a novel and flexible architecture for the UT, whereby we optimize a number of uncorrelated phase shifters groups, number of antennas per group and the set of antennas inside each group with the goal of maximizing the EE at the UT. The major novelty and contributions of this paper are as follows

- We propose novel UT architectures consisting of two stages analog beamformer in the UL and combiner in the DL. We consider an analog phase shifter stage to attain high gain and combat the severe path-loss, and apply the analog switches stage to perform antenna selection and grouping to maximize the EE at the UT.
- We exploit the spatial modulation principles to transmit two streams (spatially modulated stream and conventionally modulated stream) using a single RF chain. Specifically, in the UL, we propose an HTSM scheme and present the analytical system model followed by two detection schemes. First, we apply an ML detector and prove closed form expressions of the SE using the mutual information. After that, we propose a reduced complexity detector with two combiners (optimal and equal ratio). In DL, we propose an HRSM scheme with a reduced complexity detector that can be implemented using the energy efficient UT architecture proposed in Fig. 1 and then, we prove a closed form expression of its SE.
- We propose a reduced complexity and efficient optimization algorithm to jointly design the precoder for the UL transmission and the combiner for the DL transmission with the purpose of maximising the EE at the UT. Specifically, the proposed algorithm jointly optimises the number of uniform linear arrays (ULA) phase shifters groups, the set of selected antennas per group and the transmit powers for the spatial symbols both in UL and DL transmissions.
- We evaluate the system performance by adopting a theoretical channel model and a realistic ray-trace based channel model to validate the results in real world like scenarios. Moreover, we compare the proposed scheme with state of the art SM and hybrid precoding systems.

The rest of the paper is organized as follows. The system model is introduced in Sec. II with the assumptions and the adopted channel models. In Sec. III, the UL HTSM

¹The extension to multiple users scenario in the DL (broadcast channel) or in the UL (multiple access channel) is the topic of a forthcoming publication.

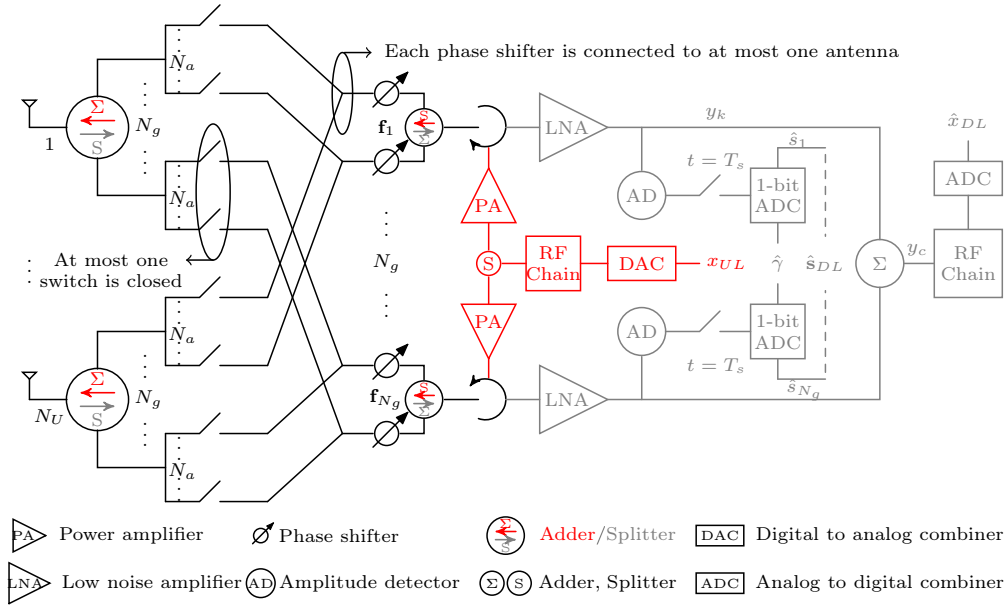


Fig. 1: Block diagram at the user terminal, with transmit and receive circuitry in red and grey respectively. Black elements (like antennas or phase shifters) are common. For a given number of active groups and phase shifters per group, the structure of the matrices \mathbf{A}_{SW} and \mathbf{A}_{PS} are detailed in tables I and II for UL and DL, respectively. Although the proposed architecture consists of single RF chain, we transmit two streams: Spatially and conventionally modulated streams.

and the DL HRSM systems are designed and the analytical SE expressions are derived. Low complexity optimization algorithms for the UL and DL systems are proposed in Sec. V. The system performance is evaluated both in stochastic and deterministic channel environments in Sec. VI. Finally, the paper is concluded in Sec. VII.

We adopt the following notations. $\mathbf{X}^{(i)}$ for the i^{th} column of \mathbf{X} , $\mathbf{X}^{(i:j)}$ for the matrix contains from i^{th} to the j^{th} columns of \mathbf{X} , $\mathbf{X}^{(i,j)}$ for the entry at the i^{th} row and j^{th} column of \mathbf{X} , $\text{Tr}\{\mathbf{X}\}$ denotes the trace operator, $\text{diag}\{\mathbf{x}\}$ denotes the diagonal matrix with elements of vector \mathbf{x} on the diagonal, \mathbf{X}^H denotes the transpose-conjugate operation and \mathbf{X}^T denotes the transpose operation.

II. SYSTEM AND CHANNEL MODELS

A. Transceiver architecture and system assumptions

The manufacturing cost and the battery lifetime of the UT are serious issues for wireless 5G modem industry. Having in mind both aspects, in Fig. 1, we propose a novel energy efficient hybrid UT architecture for the UL and DL transmissions that comprises a low number of power hungry devices (a single RF chain and a single high resolution analog-to-digital converter (ADC) regardless the number of antennas at the UT) and power efficient devices (RF amplitude detectors (AD) [27], 1-bit ADCs, switches and phase shifters). We consider a fully connected hybrid base station (BS) architecture [11] and assume a few number of RF chains that ensure the hybrid precoding/combining exactly implements a digital precoding/combining. This can be achieved if the number of RF chains is larger or equal to number of channel scattering clusters (C) [28] or at least twice the data streams [15]. The proposed hybrid architecture at the UT consists of two analog stages. The phase shifting stage provides high transmit

beamforming gain and high receive combining gain during UL and DL transmissions, respectively. The phase shifting architecture consists of N_g groups of linear antenna arrays where each group comprises N_a phase shifters. As we map the spatial bits into phase shifting groups, we consider analog switches stage to obtain uncorrelated groups. This can be achieved by smartly mapping the antennas among the phase shifting groups. Specifically, each antenna can be connected to any phase shifter and hence $N_a N_g$ ($N_a^{\text{UL}} N_g^{\text{UL}}$ for the UL and $N_a^{\text{DL}} N_g^{\text{DL}}$ for the DL architectures) switches per antenna are required. The maximum number of groups is $N_g = C$ (around 7 as shown in the realistic urban scenario channel results in Sec. VI-B) and the maximum of phase shifters per group is $N_a = N_U$. Within a specific linear antenna array group, each phase shifter is connected to a distinct antenna but different groups can share the same antenna. The number of active phase shifting groups $\{N_g^{\text{UL}}, N_g^{\text{DL}}\}$, antennas inside the groups and the number of active phase shifters per group $\{N_a^{\text{UL}}, N_a^{\text{DL}}\}$ are determined to maximize the EE at the beginning of each coherence time by employing the low complexity optimization algorithm proposed in Sec. V. The DL and the UL power consumption of the proposed UT can be expressed as

$$\begin{aligned}
 P_c^{\text{DL}} &= N_a^{\text{DL}} N_g^{\text{DL}} (P_{\text{SW}} + P_{\text{PS}}) + N_g^{\text{DL}} P_{\text{LNA}} \\
 &\quad + P_{\text{RF}} + P_{\text{ADC}} + P_{\text{BB}}, \\
 P_c^{\text{UL}} &= N_a^{\text{UL}} N_g^{\text{UL}} (P_{\text{SW}} + P_{\text{PS}}) + N_g^{\text{UL}} P_{\text{PA}} \\
 &\quad + P_{\text{RF}} + P_{\text{DAC}} + P_{\text{BB}},
 \end{aligned} \tag{1}$$

where $\{N_a^{\text{UL}} N_g^{\text{UL}}, N_a^{\text{DL}} N_g^{\text{DL}}\}$ switches are on during the coherence time in the UL and DL transmissions, respectively and the rest of switches remain off. The power consumed by UT devices at 28GHz [14]-[29] can be modelled as

$$P_{\text{PS}} = P_{\text{LNA}} = P_{\text{ref}}, P_{\text{ADC}} = P_{\text{DAC}} = \text{FoM} \times f_s \times 2^n,$$

$$\begin{aligned}
P_{\text{SW}} &= 0.25P_{\text{ref}}, P_{\text{RF}} = 2P_{\text{ref}}, P_{\text{BB}} = 10P_{\text{ref}}, \\
P_{\text{PA}} &= \left(\frac{1}{\eta} - 1\right) P_t.
\end{aligned} \tag{2}$$

Therein, P_{LNA} refers to the power consumption of the low noise amplifier (LNA) and it is taken as the reference P_{ref} in the hardware circuitry. Furthermore, the power consumption of the remaining hardware elements such as phase shifter (PS), power amplifier (PA), ADC, digital-to-analog converter (DAC), switch (SW), RF chain and baseband computation (BB) are defined by using the reference power consumption value P_{ref} , f_s is the sampling frequency that equals to twice the bandwidth, n refers to number of ADC/DAC bits, FoM is the figure of merit that depends on the technology and takes value of 34.4 fJ/Conv.-step at $n = 12$ and $f_s = 600\text{-MS/s}$ [30], η is the power amplifier efficiency that takes value 40% at 28 GHz [31] and P_t is the transmit power. The power consumption of the AD is negligible.

Each UT antenna is connected to a splitter/adder block that is activated either during the DL or the UL phase. The splitter functionality is utilized to split the received signal from the designated antenna to the phase shifters through the switches to perform analog combining at the DL receiving phase. The adder block is utilized to add the signals coming from different phase shifters to the selected antennas through the switches to perform analog beamforming during the UL transmission phase. Only those antennas that are selected as active by the optimization algorithm (see Sec. V) contribute to the communication and the rest remains idle.

The incoming data stream prior to the transmission comprises of two parts. The former is modulated according to a conventional M -ary modulation scheme, the latter is mapped onto the indices of the active antenna groups. Under the TSM principle, during the UL phase, N_g^{UL} groups of phase shifters transmit the same M^{UL} -ary modulation symbol from the activated antennas either with high power (UL spatial bit 1) or with low power (UL spatial bit 0) that results in transmitting N_g^{UL} spatial modulated bits and $\log_2 M^{\text{UL}}$ conventionally modulated bits. The BS employs ZF combiner to detect the UL spatial and modulation symbols. Similarly, in the DL phase and under the RSM principle, the BS applies ZF precoder in such a way that N_g^{DL} groups of phase shifters receive the same M^{DL} -ary modulation symbol from the activated antennas either with high or low power based on the DL spatial symbol that results in receiving N_g^{DL} spatial modulated bits and $\log_2 M^{\text{DL}}$ conventional modulated bits.

The duplexing protocol is assumed to be time-division duplex (TDD) where the CSI is needed only at the BS and the channel reciprocity is assumed. The BS can acquire the CSI during the UL training phase by any method, for instance the authors in [9] exploit the spatially sparse nature of the outdoor mmWave channel in developing low complexity adaptive compress sensing based algorithms. Imperfect CSI at the BS with ZF precoder can be accurately modelled as an increase in the noise power [32]. Thereafter, the BS runs the optimization algorithm detailed in Sec. V to determine the system parameters detailed in Sec. III and Sec. IV. The BS informs the UT about the results of the optimization algorithm

during the DL training phase. Since the information required by the UT is limited, the DL training phase results in a low training overhead. Moreover, the BS applies the ZF precoder and the combiner to employ during DL and UL transmissions, respectively using a hybrid architecture like the one proposed in [28].

B. Stochastic channel model

In the mmWave band, the number of scatterers typically assumed to be a few, as a result of the severe path-loss of the waves traveling at high frequencies. In order to take this effect into account in the system performance evaluation, we adopt a geometry-based channel model [22] whereby the UL channel matrix is given by

$$\mathbf{H} = \sqrt{\frac{N_{\text{BS}}N_U}{P_t C}} \sum_{i=1}^C g_i \mathbf{v}_r(\theta_i) \mathbf{v}_t(\phi_i)^H. \tag{3}$$

Herein, P_t is the path-loss of the channel $\mathbf{H} \in \mathbb{C}^{N_U \times N_{\text{BS}}}$ between the BS and the UT, g_i is the gain of the i -th path that follows a complex Gaussian distribution as $\mathcal{CN}(0, 1)$, $\theta_i \sim \mathcal{U}[-\pi/6, \pi/6]$ and $\phi_i \sim \mathcal{U}[-\pi/2, \pi/2]$ represent the azimuth angles of arrival at the BS and departure from the UT. By assuming ULA, transmit and receive array response vectors of the i -th path $\mathbf{v}_t(\phi_i)$ and $\mathbf{v}_r(\theta_i)$ are generated as $\mathbf{v}(\varphi) = 1/N [1, e^{jkd \sin(\varphi)}, \dots, e^{j(N-1)kd \sin(\varphi)}]^T$, where φ is the angle of the considered path, N is the number of elements in the array, $k = \frac{2\pi}{\lambda}$ where λ is the signal wavelength and $d = \frac{\lambda}{2}$ is the inter-elements spacing. The channel model in (3) can be decomposed as $\mathbf{H} = \mathbf{A}_r \mathbf{D} \mathbf{A}_t^H$ where $\mathbf{A}_r \in \mathbb{C}^{N_U \times C}$ and $\mathbf{A}_t \in \mathbb{C}^{N_{\text{BS}} \times C}$ comprise the array response vectors of all the paths $\mathbf{A}_r = [\mathbf{v}_r(\theta_1), \mathbf{v}_r(\theta_2), \dots, \mathbf{v}_r(\theta_C)]$ and $\mathbf{A}_t = [\mathbf{v}_t(\phi_1), \mathbf{v}_t(\phi_2), \dots, \mathbf{v}_t(\phi_C)]$. The diagonal matrix $\mathbf{D} \in \mathbb{C}^{C \times C}$ has the complex path gains and the path loss at the diagonal entries $\sqrt{N_{\text{BS}}N_U/P_t C} [g_1, g_2, \dots, g_C]$.

In this study, we adopt the stochastic channel model given in (3) in order to evaluate the system performance. Although the considered channel model is widely used in literature and provides analogous model for mmWave channel environment, real world channels are highly dependent on the propagation scenario. Therefore, we validate the performance of the proposed system model in Sec. VI-B with the channels predicted from the ray-based Volcano technology by SIRADEL [33].

III. UPLINK HYBRID TRANSMIT SPATIAL MODULATION

We propose a novel two stages analog precoding aided HTSM scheme assuming ZF combiner at the BS and energy efficient UT architecture. In order to combat the severe path-loss associated with the mmWave propagation, we apply a transmit beamformer stage that consists of linear arrays of phase shifters. As the application of the ZF combiner at the BS requires a full-rank MIMO channel, we use at the UT analog switches stage to perform antenna selection and grouping that reduces the correlation among the UT antennas and ensures the rank condition. Antenna selection is needed even if with MMSE combiner, as illustrated in [34]. The design of the UL transmitter is done in the following steps. First, the phase shifters stage consists of N_g^{UL} transmit analog beamformers that are employed to boost the transmit beamforming gain. Each analog beamformer contains N_a^{UL} active phase shifters.

TABLE I: Elements of the uplink signal model

x^{UL}	UL modulation symbol $\in \mathbb{C}^{1 \times 1}$ with $\mathbb{E}[x^{\text{UL}}x^{\text{UL}H}] = 1$
\mathbf{s}_i^{UL}	UL spatial symbol $\in \mathbb{R}^{N_g^{\text{UL}} \times 1}$, $i = 1, \dots, 2^{N_g^{\text{UL}}}$ mapped from N_g^{UL} bits from the incoming data bits with $\Pr(\mathbf{s}_i^{\text{UL}}) = 1/2^{N_g^{\text{UL}}}$
\mathbf{t}_i^{UL}	the mapped version of the spatial symbol \mathbf{s}_i^{UL} to high and low amplitudes $\{a_H^{\text{UL}} = 1 - a_0^{\text{UL}}, a_L^{\text{UL}} = a_0^{\text{UL}}\}$; respectively and $\{0 \leq a_0^{\text{UL}} \leq \frac{1}{2}\}$
$\mathbf{A}_{\text{PS}}^{\text{UL}}$	UL analog phase shifters matrix $\in \mathbb{C}^{N_a^{\text{UL}} \times N_g^{\text{UL}} \times N_g^{\text{UL}}}$, $\mathbf{A}_{\text{PS}}^{\text{UL}} = \text{blockdiag}\{\mathbf{f}_1^{\text{UL}}, \dots, \mathbf{f}_{N_g^{\text{UL}}}^{\text{UL}}\}$
\mathbf{f}_l^{UL}	UL analog beamformer response vector $\in \mathbb{C}^{N_a^{\text{UL}} \times 1}$
$\mathbf{A}_{\text{SW}}^{\text{UL}}$	UL analog switches matrix $\in \mathbb{R}^{N_U \times N_a^{\text{UL}} \times N_g^{\text{UL}}}$, $\mathbf{A}_{\text{SW}}^{\text{UL}(i,j)} \in \{0, 1\}$, $i = 1, \dots, N_U, j = 1, \dots, N_a^{\text{UL}} N_g^{\text{UL}}$,

Next, the switches stage is used to combat the spatial correlation among the UT antennas and as a result the received power at the BS is maximized. This is achieved by selecting the best number and the best set of the UT antennas to be connected to each transmit analog beamformer. Mapping the incoming UL bit stream to symbols is performed in two parts, the first N_g^{UL} bits (spatially modulated bits) are mapped into transmit power levels of the N_g^{UL} transmit analog beamformers such that the i^{th} beamformer transmits high or low power if the i^{th} spatial bit is 1 or 0, respectively. The remaining $\log_2 M^{\text{UL}}$ bits (modulation bits) are modulated using standard M^{UL} -ary modulation schemes. As a result, the signal transmitted to the BS can be expressed as

$$\begin{aligned} \mathbf{x}_t^{\text{UL}} &= \sqrt{\alpha^{\text{UL}}} P_t \mathbf{A}_{\text{SW}}^{\text{UL}} \mathbf{A}_{\text{PS}}^{\text{UL}} \mathbf{t}_i^{\text{UL}} x^{\text{UL}} \quad \text{with} \\ \mathbf{t}_i^{\text{UL}} &= (1 - 2a_0^{\text{UL}}) \mathbf{s}_i^{\text{UL}} + a_0^{\text{UL}} \mathbf{1}_{N_g^{\text{UL}}}, \end{aligned} \quad (4)$$

where the details of the parameters are given in Table I. Moreover, \mathbf{f}_l^{UL} is the phase shifters response vector so it has constant amplitude and satisfies $|\mathbf{f}_l^{\text{UL}(i)}| = |\mathbf{f}_l^{\text{UL}(j)}| \forall i, j, l = 1, \dots, N_g^{\text{UL}}$, $\mathbf{A}_{\text{SW}}^{\text{UL}}$ connects the phase shifters to the UT antennas in such a way that $\mathbf{A}_{\text{SW}}^{\text{UL}(i,j)} = 1$ when the j^{th} phase shifter represented by the j^{th} column of $\mathbf{A}_{\text{SW}}^{\text{UL}}$ is connected to the i^{th} UT antenna represented by the i^{th} row of $\mathbf{A}_{\text{SW}}^{\text{UL}}$ and hence, $\|\mathbf{A}_{\text{SW}}^{\text{UL}(j)}\|_0 = 1, j = 1, \dots, N_a^{\text{UL}} N_g^{\text{UL}}$. The phase shifter inside a specific beamformer is connected to distinct UT antenna and thus $\mathbf{A}_{\text{SW}}^{\text{UL}}$ satisfies $\|\mathbf{A}_{\text{SW}}^{\text{UL}(i,(k-1)N_a^{\text{UL}}+1:kN_a^{\text{UL}})}\|_0 \in \{0, 1\}, i = 1, \dots, N_U, k = 1, \dots, N_g^{\text{UL}}$. Finally, α^{UL} is the coefficient that ensures the constant average transmit power and can be expressed as

$$\begin{aligned} \alpha^{\text{UL}} &= \frac{1}{\mathbb{E}[\|\mathbf{A}_{\text{SW}}^{\text{UL}} \mathbf{A}_{\text{PS}}^{\text{UL}} \mathbf{t}_i^{\text{UL}} x^{\text{UL}}\|_2^2]} \\ &= \frac{1}{\text{Tr}\{\mathbf{A}_{\text{SW}}^{\text{UL}} \mathbf{A}_{\text{PS}}^{\text{UL}} \mathbf{R}_{ss}^{\text{UL}} \mathbf{A}_{\text{PS}}^{\text{UL}H} \mathbf{A}_{\text{SW}}^{\text{UL}H}\}}, \\ \text{with } \mathbf{R}_{ss}^{\text{UL}} &= \frac{1}{2^{N_g^{\text{UL}}}} \sum_{i=1}^{2^{N_g^{\text{UL}}}} \mathbf{t}_i^{\text{UL}} \mathbf{t}_i^{\text{UL}H}. \end{aligned} \quad (5)$$

As an illustrative example to link matrix $\mathbf{A}_{\text{SW}}^{\text{UL}}$ and the proposed architecture in Fig. 1, let us consider $N_U = 4$, $N_g^{\text{UL}} = 2$, $N_a^{\text{UL}} = 2$ and the UL switching matrix is

$$\mathbf{A}_{\text{SW}}^{\text{UL}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}, \quad (6)$$

The first row of $\mathbf{A}_{\text{SW}}^{\text{UL}}$ in Eq. (6) shows that the first antenna is connected to the first phase shifter of the first group. The third row of $\mathbf{A}_{\text{SW}}^{\text{UL}}$ shows that the third antenna is connected to the first phase shifter of the second group. Similarly, the fourth row of $\mathbf{A}_{\text{SW}}^{\text{UL}}$ describes that the fourth antenna combines the signals coming from second phase shifter of the first group and second phase shifter of the second group. Our target in Sec. V, Algorithm 2 is to jointly optimise a_0^{UL} , $\mathbf{A}_{\text{PS}}^{\text{UL}}$ and $\mathbf{A}_{\text{SW}}^{\text{UL}}$ to maximize the EE under SE requirements. The received symbol at the BS is

$$\mathbf{r}^{\text{UL}} = \sqrt{\alpha^{\text{UL}}} P_t \mathbf{H}^{\text{UL}} \mathbf{A}_{\text{SW}}^{\text{UL}} \mathbf{A}_{\text{PS}}^{\text{UL}} \mathbf{t}_i^{\text{UL}} x^{\text{UL}} + \mathbf{n}^{\text{UL}}. \quad (7)$$

Therein, $\mathbf{r}^{\text{UL}} \in \mathbb{C}^{N_{\text{BS}} \times 1}$ is the received signal vector at the BS and $\mathbf{n}^{\text{UL}} \in \mathbb{C}^{N_{\text{BS}} \times 1}$ is the noise vector with independent and identically distributed (i.i.d.) circularly symmetric complex Gaussian elements $\mathcal{CN}(0, \sigma^2)$. Furthermore, the effective UL channel matrix can be indicated as $\mathbf{H}_e^{\text{UL}} = \mathbf{H}^{\text{UL}} \mathbf{A}_{\text{SW}}^{\text{UL}} \mathbf{A}_{\text{PS}}^{\text{UL}} = [\mathbf{H}_1^{\text{UL}} \mathbf{f}_1^{\text{UL}}, \dots, \mathbf{H}_{N_g^{\text{UL}}}^{\text{UL}} \mathbf{f}_{N_g^{\text{UL}}}^{\text{UL}}] \in \mathbb{C}^{N_{\text{BS}} \times N_g^{\text{UL}}}$ where $\mathbf{H}_k^{\text{UL}} = \mathbf{H}^{\text{UL}} \mathbf{A}_{\text{SW}}^{\text{UL}((k-1)N_a^{\text{UL}}+1:kN_a^{\text{UL}})} \in \mathbb{C}^{N_{\text{BS}} \times N_a^{\text{UL}}}$ is the effective sub-channel matrix of the k^{th} beamforming group. In the UL reception, the BS applies a ZF combiner to enable the spatial and modulation symbols detection as follows

$$\mathbf{y}^{\text{UL}} = \mathbf{W} \mathbf{r}^{\text{UL}}, \quad (8)$$

where the ZF combiner matrix is computed as $\mathbf{W} = (\mathbf{H}_e^{\text{UL}H} \mathbf{H}_e^{\text{UL}})^{-1} \mathbf{H}_e^{\text{UL}H}$ and can be implemented using the hybrid architecture proposed in [28]. In Sec. V, we select the beamforming groups and the antennas per group to ensure full rank effective channel \mathbf{H}_e^{UL} . The post-processed signal $\mathbf{y}^{\text{UL}} \in \mathbb{C}^{N_g^{\text{UL}} \times 1}$ comprises the spatial and modulation symbols. Hence, the k^{th} entry of \mathbf{y}^{UL} takes the following values

$$y_k^{\text{UL}} = \begin{cases} \sqrt{\alpha^{\text{UL}}} P_t a_H^{\text{UL}} x^{\text{UL}} + n'_k & \text{if } s_{ik}^{\text{UL}} = 1 \\ \sqrt{\alpha^{\text{UL}}} P_t a_L^{\text{UL}} x^{\text{UL}} + n'_k & \text{if } s_{ik}^{\text{UL}} = 0 \end{cases} \quad (9)$$

where $n'_k \in \mathcal{CN}(0, \sigma_k'^2)$ is the k^{th} entry of the post-processed noise variable $\mathbf{n} = \mathbf{W} \mathbf{n}^{\text{UL}}$ with variance $\sigma_k'^2 = [\mathbf{W} \mathbf{W}^H]^{(k,k)} \sigma^2$.

A. Uplink maximum likelihood detector

Since we assume that the number of RF chains at the BS is at least C , we can jointly detect the spatial and the modulation symbols using an ML approach as:

$$[\hat{\mathbf{s}}^{\text{UL}}, \hat{x}^{\text{UL}}] = \max_{\mathbf{s}_i^{\text{UL}}, x_j^{\text{UL}}} f(\mathbf{y}^{\text{UL}} | \mathbf{s}_i^{\text{UL}}, x_j^{\text{UL}})$$

$$= \min_{\mathbf{s}_i^{\text{UL}}, x_j^{\text{UL}}} \left\| \mathbf{R}_{n'n'}^{-\frac{1}{2}} \left(\mathbf{y}^{\text{UL}} - \sqrt{\alpha^{\text{UL}}} \mathbf{t}_i^{\text{UL}} x_j^{\text{UL}} \right) \right\|_2^2, \quad (10)$$

where $\mathbf{R}_{n'n'} = \sigma^2 \mathbf{W} \mathbf{W}^H$. Although the ML detector provides optimal performance, exhaustive search in Eq. (10) is computationally complex if a large number of bits per spatial symbol is being transmitted. Since our goal is to maximize the EE under SE constraint, we prove in the sequel an expression for the SE of the ML detector assuming Gaussian x^{UL} , and propose a low-complexity detector assuming M -PSK x^{UL} . Finally, we show that the SE of the reduced complexity detector achieves a tight lower bound on the mutual information.

B. Spectral efficiency of the maximum likelihood detector with Gaussian x^{UL}

metric to evaluate the proposed scheme. By applying the mutual information chain rule, the SE can be defined as

$$\begin{aligned} I(\mathbf{s}^{\text{UL}}, x^{\text{UL}}; \mathbf{y}^{\text{UL}}) &= I(\mathbf{s}^{\text{UL}}; \mathbf{y}^{\text{UL}}) + I(x^{\text{UL}}; \mathbf{y}^{\text{UL}} | \mathbf{s}^{\text{UL}}), \\ I(\mathbf{s}^{\text{UL}}; \mathbf{y}^{\text{UL}}) &= h(\mathbf{y}^{\text{UL}}) - h(\mathbf{y}^{\text{UL}} | \mathbf{s}^{\text{UL}}) \triangleq I_S^{\text{UL}}, \\ I(x^{\text{UL}}; \mathbf{y}^{\text{UL}} | \mathbf{s}^{\text{UL}}) &= h(\mathbf{y}^{\text{UL}} | \mathbf{s}^{\text{UL}}) - h(\mathbf{y}^{\text{UL}} | \mathbf{s}^{\text{UL}}, x^{\text{UL}}) \triangleq I_M^{\text{UL}}. \end{aligned} \quad (11)$$

Assuming Gaussian-distributed x^{UL} , the differential entropies in Eq. (11) can be computed from

$$h(\mathbf{y}^{\text{UL}}) = - \int_{\mathcal{C}^{N_g^{\text{UL}}}} f(\mathbf{y}^{\text{UL}}) \log_2 f(\mathbf{y}^{\text{UL}}) d\mathbf{y}^{\text{UL}} \quad (12)$$

where $f(\mathbf{y}^{\text{UL}})$ is the probability density function (PDF) of the complex Gaussian mixture random vector

$$\begin{aligned} f(\mathbf{y}^{\text{UL}}) &= \sum_{i=1}^{2^{N_g^{\text{UL}}}} \Pr(\mathbf{s}_i^{\text{UL}}) f(\mathbf{y}^{\text{UL}} | \mathbf{s}_i^{\text{UL}}) \\ &= \frac{1}{2^{N_g^{\text{UL}}}} \sum_{i=1}^{2^{N_g^{\text{UL}}}} \frac{1}{\pi^{N_g^{\text{UL}}} |\boldsymbol{\Sigma}_i|} e^{-\mathbf{y}^{\text{UL}H} \boldsymbol{\Sigma}_i^{-1} \mathbf{y}^{\text{UL}}} \quad \text{with} \\ \boldsymbol{\Sigma}_i &= \alpha^{\text{UL}} P_t \mathbf{t}_i^{\text{UL}H} \mathbf{t}_i^{\text{UL}} + \mathbf{R}_{n'n'}. \end{aligned} \quad (13)$$

The closed form expression of the differential entropy of a Gaussian mixture $h(\mathbf{y}^{\text{UL}})$ is unknown [35]. Moreover, the numerical evaluation of $h(\mathbf{y}^{\text{UL}})$ is computationally complex especially in large-scale MIMO systems when the size of \mathbf{y}^{UL} can be large. In the sequel, we propose a novel and reduced complexity method to evaluate $h(\mathbf{y}^{\text{UL}})$. In this method, we apply the conditional entropy chain rule on $h(\mathbf{y}^{\text{UL}})$ and we prove a closed form expression for the conditional probability density function $f(y_k^{\text{UL}} | y_{k-1}^{\text{UL}}, \dots, y_1^{\text{UL}})$. Finally, we reduce the computational complexity by simplifying the integral in Eq. (12) to be sum of double integrals as follows

$$\begin{aligned} h(\mathbf{y}^{\text{UL}}) &= h(y_1^{\text{UL}}) + \sum_{k=2}^{N_g^{\text{UL}}} h(y_k^{\text{UL}} | y_{k-1}^{\text{UL}}, \dots, y_1^{\text{UL}}) \quad \text{with} \\ f(y_k^{\text{UL}} | y_{k-1}^{\text{UL}}, \dots, y_1^{\text{UL}}) &= \\ \frac{1}{2^k} \sum_{i=1}^{2^k} f(y_k^{\text{UL}} | y_{k-1}^{\text{UL}}, \dots, y_1^{\text{UL}}, \mathbf{t}_i^{\text{UL}}(1:k)). \end{aligned} \quad (14)$$

Lemma 1. *The conditional density function of the random variable $(y_k^{\text{UL}} | y_{k-1}^{\text{UL}}, \dots, y_1^{\text{UL}}, \mathbf{t}_i^{\text{UL}}(1:k))$ is distributed as a zero-mean complex Gaussian :*

$$f(y_k^{\text{UL}} | y_{k-1}^{\text{UL}}, \dots, y_1^{\text{UL}}, \mathbf{t}_i^{\text{UL}}(1:k)) = \mathcal{CN}(0, \sigma_K^2 + P_{i,K}).$$

The proof of Lemma 1 and the values of σ_K^2 and $P_{i,K}$ are in Appendix A.

Therefore, we can determine the k^{th} differential entropy $h(y_k^{\text{UL}} | y_{k-1}^{\text{UL}}, \dots, y_1^{\text{UL}})$ by evaluating the double integral numerically regardless the size of \mathbf{y}^{UL} . Thus, we significantly reduce the computational complexity. As an illustrative example that highlights the proposed reduced complexity method of evaluating $h(\mathbf{y}^{\text{UL}})$, let us consider $N_g^{\text{UL}} = 3$. The proposed 2-D integral method achieves the same exact values of $h(\mathbf{y}^{\text{UL}})$ as the 16-D integral in Eq. (12). The differential entropy $h(\mathbf{y}^{\text{UL}} | \mathbf{s}^{\text{UL}})$ can be expressed in closed form as

$$\begin{aligned} h(\mathbf{y}^{\text{UL}} | \mathbf{s}^{\text{UL}}) &= \sum_{i=1}^{2^{N_g^{\text{UL}}}} \Pr(\mathbf{s}_i^{\text{UL}}) h(\mathbf{y}^{\text{UL}} | \mathbf{s}^{\text{UL}} = \mathbf{s}_i^{\text{UL}}) \\ &= \frac{1}{2^{N_g^{\text{UL}}}} \sum_{i=1}^{2^{N_g^{\text{UL}}}} \log_2(\pi e) \left| \alpha^{\text{UL}} P_t \mathbf{t}_i^{\text{UL}H} \mathbf{t}_i^{\text{UL}} + \mathbf{R}_{n'n'} \right| \\ &= \frac{1}{2^{N_g^{\text{UL}}}} \sum_{i=1}^{2^{N_g^{\text{UL}}}} \left(N_g^{\text{UL}} \log_2(\pi e) \right. \\ &\quad \left. + \log_2 |\mathbf{R}_{n'n'}| \left(1 + \alpha^{\text{UL}} P_t \mathbf{t}_i^{\text{UL}H} \mathbf{R}_{n'n'}^{-1} \mathbf{t}_i^{\text{UL}} \right) \right). \end{aligned} \quad (15)$$

Based on the low computational complexity method of evaluating $h(\mathbf{y}^{\text{UL}})$ and the closed form of $h(\mathbf{y}^{\text{UL}} | \mathbf{s}^{\text{UL}})$, we can evaluate the spatial rate I_S^{UL} in Eq. (11) efficiently. Moreover, the differential entropy $h(\mathbf{y}^{\text{UL}} | \mathbf{s}^{\text{UL}}, x)$ can be expressed as

$$h(\mathbf{y}^{\text{UL}} | \mathbf{s}^{\text{UL}}, x) = h(\mathbf{n}') = \log_2(\pi e) N_g^{\text{UL}} |\mathbf{R}_{n'n'}|. \quad (16)$$

According to Eq. (15) and Eq. (16), I_M^{UL} in Eq. (11) can be expressed in closed form as

$$I_M^{\text{UL}} = \frac{1}{2^{N_g^{\text{UL}}}} \sum_{i=1}^{2^{N_g^{\text{UL}}}} \log_2 \left(1 + \alpha^{\text{UL}} P_t \mathbf{t}_i^{\text{UL}H} \mathbf{R}_{n'n'}^{-1} \mathbf{t}_i^{\text{UL}} \right). \quad (17)$$

Finally, according to Eq. (11), the UL SE can be evaluated as $\text{SE}^{\text{UL}} = I_S^{\text{UL}} + I_M^{\text{UL}}$.

C. Uplink reduced complexity detection with M -PSK x^{UL}

Let us propose a low complexity detection method whereby the spatial and the modulation symbols are detected separately. The size of the search space for the reduced complexity detector is $N_a^{\text{UL}} + M$ which is much smaller than that of the ML detector ($2^{N_a^{\text{UL}}} \times M$). In Fig. 2 in [36], the authors showed that constant amplitude constellations (M -PSK) achieves the best performance with the reduced complexity SM detector.

From Eq. (10), the ML spatial symbol detector is $\hat{\mathbf{t}}^{\text{UL}} = x^{\text{UL}} \mathbf{y}^{\text{UL}} / \sqrt{\alpha^{\text{UL}} P_t}$. However, x^{UL} is unknown so we exploit the fact that M -PSK modulation symbols have constant amplitude and hence, $|x^{\text{UL}}| = 1$. Then, in order to detect the k^{th} binary

spatially modulated bit, we compare the absolute value of y_k^{UL} with a threshold γ^{UL} as follows

$$\hat{s}_{ik}^{\text{UL}} = \begin{cases} 1 & \text{if } |y_k^{\text{UL}}| \geq \gamma^{\text{UL}} \\ 0 & \text{if } |y_k^{\text{UL}}| < \gamma^{\text{UL}} \end{cases} \quad \text{with} \\ \gamma^{\text{UL}} = \frac{1}{2} \sqrt{\alpha^{\text{UL}} P_t} (a_H^{\text{UL}} + a_L^{\text{UL}}). \quad (18)$$

According to Eq. (10), the optimal modulation symbol detector can be expressed as

$$x_j^{\text{UL}*} = \max_{x_j^{\text{UL}} \in M\text{-PSK}} \Re \left\{ x_j^{\text{UL}H} \mathbf{t}_i^{\text{UL}H} \mathbf{R}_{n',n}^{-1} \mathbf{y}^{\text{UL}} \right\} \\ = \min_{x_j^{\text{UL}} \in M\text{-PSK}} |x_j^{\text{UL}} - \mathbf{v}_i^{\text{OCH}} \mathbf{y}^{\text{UL}}|_2 \quad \text{with} \\ \mathbf{v}_i^{\text{OC}} = \mathbf{R}_{n',n}^{-1} \mathbf{t}_i^{\text{UL}} = \mathbf{H}_e^{\text{UL}H} \mathbf{H}_e^{\text{UL}} \mathbf{t}_i^{\text{UL}}, \quad (19)$$

where \mathbf{v}_i^{OC} is the optimal modulation symbol combiner. Since \mathbf{t}_i^{UL} is unknown at the UT. We propose to use the detected spatial symbol in Eq. (18). Therefore, the combined signal can be expressed as

$$y_c^{\text{UL,OC}} = \sqrt{\alpha^{\text{UL}} P_t} \hat{\mathbf{t}}_i^{\text{UL}T} \mathbf{H}_e^{\text{UL}H} \mathbf{H}_e^{\text{UL}} \mathbf{t}_i^{\text{UL}} x^{\text{UL}} \\ + \hat{\mathbf{t}}_i^{\text{UL}T} \mathbf{H}_e^{\text{UL}H} \mathbf{n}^{\text{UL}}. \quad (20)$$

Eq. (20) shows that the SNR perceived for the spatial symbol detection affects the modulation symbol detection. Thus, we propose the use of an equal ratio combiner (ERC) combined signal that is independent from the detected spatial symbol as

$$y_c^{\text{UL,ERC}} = \mathbf{1}_{N_g^{\text{UL}}}^T \mathbf{y}^{\text{UL}} = \sqrt{\alpha^{\text{UL}} P_t} \mathbf{1}_{N_g^{\text{UL}}}^T \mathbf{t}_i^{\text{UL}} x^{\text{UL}} \\ + \mathbf{1}_{N_g^{\text{UL}}}^T \mathbf{W} \mathbf{n}^{\text{UL}}. \quad (21)$$

After that, we apply minimum distance detector on the combined signal to detect the M -PSK symbol. In the sequel, we provide closed form expressions for the UL SE for both combiners.

D. Spectral efficiency of the reduced complexity detector with M -PSK x^{UL}

As we detect the spatial and the modulation symbols independently, the SE of the proposed reduced complexity detection UL HTSM scheme can be expressed as

$$\text{SE}_r^{\text{UL}} = I_{S,r}^{\text{UL}} + I_{M,r}^{\text{UL}}. \quad (22)$$

We transmit binary spatial symbol with input $s_{ik}^{\text{UL}} \in \{0, 1\}$, $k = 1, \dots, N_g^{\text{UL}}$ and output $\hat{s}_{ik}^{\text{UL}} \in \{0, 1\}$ as in Eq. (18) and thus, the wireless channel between s_{ik}^{UL} and \hat{s}_{ik}^{UL} can be characterized by the binary asymmetric channel (BAC) [37]. Hence, the UL spatial rate (defined as $I_{S,r}^{\text{UL}}$) can be expressed as a contribution of N_g^{UL} parallel BACs as

$$I_{S,r}^{\text{UL}} = I(\mathbf{s}^{\text{UL}}, \hat{\mathbf{s}}^{\text{UL}}) = \sum_{k=1}^{N_g^{\text{UL}}} I(s^{\text{UL}}, \hat{s}^{\text{UL}}) \\ = \sum_{k=1}^{N_g^{\text{UL}}} \mathcal{H} \left(\frac{P_{0k} + 1 - P_{1k}}{2} \right) - \frac{\mathcal{H}(P_{0k}) + \mathcal{H}(1 - P_{1k})}{2}. \quad (23)$$

Herein, the entropy function is $\mathcal{H}(p) = -p \log_2(p) - (1-p) \log_2(1-p)$ and the probabilities of the false detection of the spatially modulated bits 1 and 0 are

$$P_{1k} = \Pr(\hat{s}_{ik}^{\text{UL}} = 1 | s_{ik}^{\text{UL}} = 0) = \Pr(|y_k^{\text{UL}}| > \gamma^{\text{UL}} | s_{ik}^{\text{UL}} = 0) \\ = 1 - Q_1 \left(\sqrt{2\alpha^{\text{UL}} P_t} a_H^{\text{UL}} / \sigma'_k, \sqrt{2}\gamma^{\text{UL}} / \sigma'_k \right), \quad (24) \\ P_{0k} = \Pr(\hat{s}_{ik}^{\text{UL}} = 0 | s_{ik}^{\text{UL}} = 1) = \Pr(|y_k^{\text{UL}}| < \gamma^{\text{UL}} | s_{ik}^{\text{UL}} = 1) \\ = Q_1 \left(\sqrt{2\alpha^{\text{UL}} P_t} a_L^{\text{UL}} / \sigma'_k, \sqrt{2}\gamma^{\text{UL}} / \sigma'_k \right), \quad (25)$$

where $|y_k^{\text{UL}}|$ is Ricean distributed and $Q_1(\cdot)$ is the first order Marcum-Q-function.

The combined signals in Eq. (20) and in Eq. (21) include one M -PSK symbol. Thus, the UL modulation rate with OC ($I_{M,r}^{\text{UL,OC}}$) and with ERC ($I_{M,r}^{\text{UL,ERC}}$) can be expressed by the MISO rate expression with the asymptotic M -PSK shaping loss approximation [38], [39]

$$I_{M,r}^{\text{UL,OC}} = I(x^{\text{UL}}; y_c^{\text{UL,OC}} | \hat{\mathbf{s}}^{\text{UL}}, \mathbf{s}^{\text{UL}}) \\ = \sum_{i=1}^{2^{N_g^{\text{UL}}}} \Pr(\mathbf{s}_i^{\text{UL}}) \sum_{l=1}^{2^{N_g^{\text{UL}}}} \Pr(\hat{\mathbf{s}}_l^{\text{UL}} | \mathbf{s}_i^{\text{UL}}) \frac{1}{2} \log_2 \left(\frac{4\pi}{e} \text{SNR}_{i,l}^{\text{UL}} \right) \\ \text{where } \text{SNR}_{i,l}^{\text{UL}} = \frac{|\hat{\mathbf{t}}_l^{\text{UL}T} \mathbf{H}_e^{\text{UL}H} \mathbf{H}_e^{\text{UL}} \mathbf{t}_i^{\text{UL}}|^2 \alpha^{\text{UL}} P_t}{\|\hat{\mathbf{t}}_l^{\text{UL}T} \mathbf{H}_e^{\text{UL}H} \mathbf{n}^{\text{UL}}\|_2^2 \sigma^2}, \\ \Pr(\hat{\mathbf{s}}_l^{\text{UL}} | \mathbf{s}_i^{\text{UL}}) = \prod_{k=1}^{N_g^{\text{UL}}} \Pr \left(|y_k^{\text{UL}}| \begin{matrix} \hat{s}_{ik}^{\text{UL}}=1 \\ \geq \\ \hat{s}_{ik}^{\text{UL}}=0 \end{matrix} \gamma_k^{\text{UL}} | s_{ik}^{\text{UL}} \right), \\ \text{and } \Pr(\mathbf{s}_i^{\text{UL}}) = \frac{1}{2^{N_g^{\text{UL}}}}. \quad (26)$$

$$I_{M,r}^{\text{UL,ERC}} = I(x^{\text{UL}}; y_c^{\text{UL,ERC}} | \mathbf{s}^{\text{UL}}) \\ = \sum_{i=1}^{2^{N_g^{\text{UL}}}} \Pr(\mathbf{s}_i^{\text{UL}}) \frac{1}{2} \log_2 \left(\frac{4\pi}{e} \text{SNR}_i^{\text{UL}} \right) \quad \text{where} \\ \text{SNR}_i^{\text{UL}} = \frac{|\mathbf{1}_{N_g^{\text{UL}}}^T \mathbf{t}_i^{\text{UL}}|^2 \alpha^{\text{UL}} P_t}{\|\mathbf{1}_{N_g^{\text{UL}}}^T \mathbf{W} \mathbf{n}^{\text{UL}}\|_2^2 \sigma^2}. \quad (27)$$

According to Eq. (23), for higher spatial rates, the difference between the amplitude levels ($a_H^{\text{UL}}, a_L^{\text{UL}}$) should be maximized. In contrast, the signal-to-noise-ratio (SNR) of the modulation symbol shown in Eq. (26) and in Eq. (27) increases with both a_H^{UL} and a_L^{UL} . Thus, in Algorithm 2 in Sec. V, we optimise the values of a_H^{UL} and a_L^{UL} to maximize the sum of spatial plus modulation rates. To conclude the evaluation of performance for the UL, let us compute the mutual information between inputs ($\mathbf{s}^{\text{UL}}, x^{\text{UL}}$) and the outputs ($\hat{\mathbf{s}}^{\text{UL}}, y_c^{\text{UL,ERC}}$) and show that the proposed low complexity detection scheme with ERC achieves a tight lower bound on the mutual information as follows

$$I(\mathbf{s}^{\text{UL}}, x^{\text{UL}}; \hat{\mathbf{s}}^{\text{UL}}, y_c^{\text{UL,ERC}}) = I(\mathbf{s}^{\text{UL}}, x^{\text{UL}}; \hat{\mathbf{s}}^{\text{UL}}) \\ + I(\mathbf{s}^{\text{UL}}, x^{\text{UL}}; y_c^{\text{UL,ERC}} | \hat{\mathbf{s}}^{\text{UL}}) \quad \text{with} \\ I(\mathbf{s}^{\text{UL}}, x^{\text{UL}}; \hat{\mathbf{s}}^{\text{UL}}) = I(\mathbf{s}^{\text{UL}}; \hat{\mathbf{s}}^{\text{UL}}) + I(x^{\text{UL}}; \hat{\mathbf{s}}^{\text{UL}} | \mathbf{s}^{\text{UL}}), \\ I(\mathbf{s}^{\text{UL}}, x^{\text{UL}}; y_c^{\text{UL,ERC}} | \hat{\mathbf{s}}^{\text{UL}}) = I(\mathbf{s}^{\text{UL}}; y_c^{\text{UL,ERC}} | \hat{\mathbf{s}}^{\text{UL}}) \\ + I(x^{\text{UL}}; y_c^{\text{UL,ERC}} | \hat{\mathbf{s}}^{\text{UL}}, \mathbf{s}^{\text{UL}}). \quad (28)$$

Note that, since x^{UL} belongs to the constant amplitude constellation (M -PSK), the received power levels do not depend on x^{UL} . Thus, the detected spatial symbol $\hat{\mathbf{s}}^{\text{UL}}$ and x^{UL} are independent and they are also independent given \mathbf{s}^{UL} , this implies $I(x^{\text{UL}}; \hat{\mathbf{s}}^{\text{UL}} | \mathbf{s}^{\text{UL}}) = 0$. Eq. (21) shows that $y_c^{\text{UL,ERC}}$ does not depend on $\hat{\mathbf{s}}^{\text{UL}}$ and therefore, $I(x^{\text{UL}}; y_c^{\text{UL,ERC}} | \hat{\mathbf{s}}^{\text{UL}}, \mathbf{s}^{\text{UL}}) = I(x^{\text{UL}}; y_c^{\text{UL,ERC}} | \mathbf{s}^{\text{UL}})$ and $I(\mathbf{s}^{\text{UL}}; y_c^{\text{UL,ERC}} | \hat{\mathbf{s}}^{\text{UL}}) = I(\mathbf{s}^{\text{UL}}; y_c^{\text{UL,ERC}})$. Finally, Eq. (28) can be simplified as follows

$$\begin{aligned} I(\mathbf{s}^{\text{UL}}, x^{\text{UL}}; \hat{\mathbf{s}}^{\text{UL}}, y_c^{\text{UL,ERC}}) &= I(x^{\text{UL}}; y_c^{\text{UL,ERC}} | \mathbf{s}^{\text{UL}}) \\ &\quad + I(\mathbf{s}^{\text{UL}}; \hat{\mathbf{s}}^{\text{UL}}) + I(\mathbf{s}^{\text{UL}}; y_c^{\text{UL,ERC}}), \\ &= \text{SE}_r^{\text{UL}} + I(\mathbf{s}^{\text{UL}}; y_c^{\text{UL,ERC}}). \end{aligned} \quad (29)$$

Therefore, the SE in Eq. (22) is a lower bound on the mutual information with a gap ratio

$$L_g = I(\mathbf{s}^{\text{UL}}; y_c^{\text{UL,ERC}}) / \text{SE}_r^{\text{UL}}. \quad (30)$$

The mutual information $I(\mathbf{s}^{\text{UL}}; y_c^{\text{UL,ERC}})$ can be bounded as

$$\begin{aligned} I(\mathbf{s}^{\text{UL}}; y_c^{\text{UL,ERC}}) &= h(y_c^{\text{UL,ERC}}) - h(y_c^{\text{UL,ERC}} | \mathbf{s}^{\text{UL}}) \quad \text{with} \\ h(y_c^{\text{UL,ERC}}) &\leq h(y_c^{\text{UL,ERC}} | \mathbf{s}^{\text{UL}} = \mathbf{1}_{N_g^{\text{UL}}}) \\ &\stackrel{(a)}{=} \frac{1}{2} \log_2 (4\pi^3 e \max(P_i) \sigma_{c,\text{ERC}}^2) \quad \text{and} \\ h(y_c^{\text{UL,ERC}} | \mathbf{s}^{\text{UL}}) &= \frac{1}{2^{N_g^{\text{UL}}}} \sum_{i=1}^{2^{N_g^{\text{UL}}}} h(y_c^{\text{UL,ERC}} | \mathbf{s}^{\text{UL}} = \mathbf{s}_i^{\text{UL}}) \\ &= \frac{1}{2^{N_g^{\text{UL}}}} \sum_{i=1}^{2^{N_g^{\text{UL}}}} \frac{1}{2} \log_2 (4\pi^3 e P_i \sigma_{c,\text{ERC}}^2) \\ I(\mathbf{s}^{\text{UL}}; y_c^{\text{UL,ERC}}) &\leq \frac{1}{2} \log_2 \frac{\max(P_i)}{2^{N_g^{\text{UL}}} \sqrt{\prod_{i=1}^{2^{N_g^{\text{UL}}}} P_i}}. \end{aligned} \quad (31)$$

Therein, $P_i = |\mathbf{1}_{N_g^{\text{UL}}}^{\text{UL}T} \mathbf{t}_i^{\text{UL}}|^2 \alpha^{\text{UL}} P_t$, $\sigma_{c,\text{ERC}}^2 = \|\mathbf{1}_{N_g^{\text{UL}}}^{\text{UL}T} \mathbf{W}\|_2^2 \sigma^2$, $\max(\cdot)$ denotes the maximum operator and step $\stackrel{(a)}{=}$ follows from the shaping loss of the M -PSK symbols [38], [39]. As an illustrative example, let us consider $\sigma_{c,\text{ERC}}^2 = 1$, $a_H^{\text{UL}} = 1$ and $a_L^{\text{UL}} = \frac{1}{2}$. Hence, Fig. 2 shows that the upper bound of the gap ratio is in range of $0.05\% \leq L_g \leq 0.4\%$ and decreases with the transmit power at different number of groups and wide range of received powers. Thus, the SE_r^{UL} of the reduced complexity detection scheme achieves a tight lower bound on the mutual information.

IV. DL HYBRID RECEIVE SPATIAL MODULATION

In the DL transmission phase, we propose a novel two stages analog combining aided HRSM scheme at the UT, assuming ZF precoding at the BS. First, a switches stage is used to select the UT antennas and connect them to the phase shifters arrays to ensure full rank equivalent channel and thus, enable the ZF precoding at the BS. Next, N_g^{DL} receive analog beamformers, each containing N_a^{DL} active phase shifters, are used to enhance the receive beamforming gain. We apply the ZF precoder at the BS on the effective channel (MIMO channel + DL RF combiner at the UT). Thus, smart design of the analog

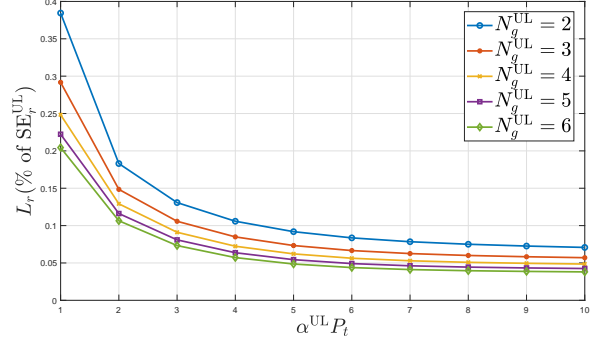


Fig. 2: Upper bound of the gap ratio of the SE_r^{UL} at different values of N_g^{UL} .

combiner at the UT boosts the received power at the UT and enhances SE. Similarly to what was done for the UL, the incoming DL bit stream is mapped into two streams. The first N_g^{DL} spatially modulated bits are mapped to the received power levels from the N_g^{DL} receive analog combiners such that the i^{th} combiner receives high or low power if the i^{th} spatial bit is 1 or 0, respectively. The remaining bits are mapped to an M -ary constellation. In this way, the BS transmits the following signal

$$\begin{aligned} \mathbf{x}_t^{\text{DL}} &= \sqrt{\alpha^{\text{DL}} P_t} \mathbf{P} \mathbf{t}_i^{\text{DL}} x^{\text{DL}} \quad \text{with} \\ \mathbf{t}_i^{\text{DL}} &= (1 - 2a_0^{\text{DL}}) \mathbf{s}_i^{\text{DL}} + a_0^{\text{DL}} \mathbf{1}_{N_g^{\text{DL}}} \quad \text{and} \\ \mathbf{t}_{ik}^{\text{DL}} &= \begin{cases} a_H^{\text{DL}} = 1 - a_0^{\text{DL}} & \text{if } s_{ik}^{\text{DL}} = 1 \\ a_L^{\text{DL}} = a_0^{\text{DL}} & \text{if } s_{ik}^{\text{DL}} = 0 \end{cases} \end{aligned} \quad (32)$$

where $x^{\text{DL}} \in \mathbb{C}^{1 \times 1}$ is the DL modulation symbol, $\{\mathbf{s}_i^{\text{DL}} \in \mathbb{R}^{N_g^{\text{DL}} \times 1}, i = 1, \dots, 2^{N_g^{\text{DL}}}\}$ is the DL spatial symbol that includes N_g^{DL} data bits and $\{\mathbf{t}_i^{\text{DL}} \in \mathbb{R}^{N_g^{\text{DL}} \times 1}, i = 1, \dots, 2^{N_g^{\text{DL}}}\}$ is the mapped version of the spatial symbol \mathbf{s}_i^{DL} . Moreover, $\mathbf{P} \in \mathbb{R}^{N_t \times N_g^{\text{DL}}}$ is the ZF precoder and α^{DL} is a normalization coefficient that fixes the average transmit power

$$\begin{aligned} \alpha^{\text{DL}} &= \frac{1}{\mathbb{E}[\|\mathbf{P} \mathbf{t}_i^{\text{DL}} x^{\text{DL}}\|_2^2]} = \frac{1}{\text{Tr}\{\mathbf{P} \mathbf{R}_{ss}^{\text{DL}} \mathbf{P}^H\}} \quad \text{with} \\ \mathbf{R}_{ss}^{\text{DL}} &= \frac{1}{2^{N_g^{\text{DL}}}} \sum_{i=1}^{2^{N_g^{\text{DL}}}} \mathbf{t}_i^{\text{DL}} \mathbf{t}_i^{\text{DL}H}. \end{aligned} \quad (33)$$

The received vector at the UT can be expressed as

$$\mathbf{r}^{\text{DL}} = \sqrt{\alpha^{\text{DL}} P_t} \mathbf{H} \mathbf{P} \mathbf{t}_i^{\text{DL}} x^{\text{DL}} + \mathbf{n}^{\text{DL}} \quad (34)$$

where \mathbf{H} is the DL channel matrix which, assuming channel reciprocity, is the transpose of the UL channel matrix $\mathbf{H} = (\mathbf{H}^{\text{UL}})^T$ and $\mathbf{n}^{\text{DL}} \in \mathbb{C}^{N_u \times 1}$ is the noise vector with i.i.d. circularly symmetric complex Gaussian elements $\mathcal{CN}(0, \sigma^2)$. At the UT, we apply a two stages analog combiner on the received vector \mathbf{r}^{DL} as follows

$$\begin{aligned} \mathbf{y}^{\text{DL}} &= \sqrt{\alpha^{\text{DL}} P_t} \mathbf{A}_{\text{PS}}^{\text{DLH}} \mathbf{A}_{\text{SW}}^{\text{DLH}} \mathbf{H} \mathbf{P} \mathbf{t}_i^{\text{DL}} x^{\text{DL}} \\ &\quad + \mathbf{A}_{\text{PS}}^{\text{DLH}} \mathbf{A}_{\text{SW}}^{\text{DLH}} \mathbf{n}^{\text{DL}} \end{aligned} \quad (35)$$

where variables are defined in Table II. Moreover, \mathbf{f}_l^{DL} has constant amplitude and thus $|\mathbf{f}_l^{\text{DL}(i)}| = |\mathbf{f}_l^{\text{DL}(j)}| = 1 \forall i, j, l = 1, \dots, N_g^{\text{DL}}$, $\mathbf{A}_{\text{SW}}^{\text{DL}}$ connects the phase shifters to the UT

TABLE II: Elements of the downlink signal model

$\mathbf{A}_{\text{PS}}^{\text{DL}}$	DL analog phase shifters matrix $\in \mathbb{C}^{N_a^{\text{DL}} N_g^{\text{DL}} \times N_g^{\text{DL}}}$, $\mathbf{A}_{\text{PS}}^{\text{DL}} = \text{blockdiag}\{\mathbf{f}_1^{\text{DL}}, \dots, \mathbf{f}_{N_g^{\text{DL}}}^{\text{DL}}\}$
\mathbf{f}_t^{DL}	DL analog combining response vector $\in \mathbb{C}^{N_a^{\text{DL}} \times 1}$
$\mathbf{A}_{\text{SW}}^{\text{DL}}$	DL analog switches matrix $\in \mathbb{R}^{N_U \times N_a^{\text{DL}} N_g^{\text{DL}}}$, $\mathbf{A}_{\text{SW}}^{\text{DL}(i,j)} \in \{0, 1\}$, $i = 1, \dots, N_U, j = 1, \dots, N_a^{\text{DL}} N_g^{\text{DL}}$

antennas such that $\mathbf{A}_{\text{SW}}^{\text{DL}(i,j)} = 1$ means that the j^{th} phase shifter represented by the j^{th} column of $\mathbf{A}_{\text{SW}}^{\text{DL}}$ is connected to the i^{th} UT antenna represented by the i^{th} row of $\mathbf{A}_{\text{SW}}^{\text{DL}}$ so $\|\mathbf{A}_{\text{SW}}^{\text{DL}(j)}\|_0 = j = 1, \dots, N_a^{\text{DL}} N_g^{\text{DL}}$, the phase shifter inside certain combiner is connected to specific UT antenna and thus $\|\mathbf{A}_{\text{SW}}^{\text{DL}(i,(k-1)N_a^{\text{DL}}+1:kN_a^{\text{DL}})}\|_0 \in \{0, 1\}$, $i = 1, \dots, N_U, k = 1, \dots, N_g^{\text{DL}}$. The received signal after the combining in Eq. (35) can be cast as

$$\begin{aligned} \mathbf{y}^{\text{DL}} &= \sqrt{\alpha^{\text{DL}}} \mathbf{P}_t \mathbf{H}_e^{\text{DL}} \mathbf{P} \mathbf{t}_t^{\text{DL}} x^{\text{DL}} + \mathbf{n}'' \text{ with} \\ \mathbf{H}_e^{\text{DL}} &= \mathbf{A}_{\text{PS}}^{\text{DLH}} \mathbf{A}_{\text{SW}}^{\text{DLH}} \mathbf{H} = \left[\mathbf{f}_1^{\text{DLH}} \mathbf{H}_1^{\text{DL}}; \dots; \mathbf{f}_{N_g^{\text{DL}}}^{\text{DLH}} \mathbf{H}_{N_g^{\text{DL}}}^{\text{DL}} \right], \\ \mathbf{n}'' &= \mathbf{A}_{\text{PS}}^{\text{DLH}} \mathbf{A}_{\text{SW}}^{\text{DLH}} \mathbf{n}^{\text{DL}}, \\ \mathbf{H}_k^{\text{DL}} &= \mathbf{A}_{\text{SW}}^{\text{DLH}((k-1)N_a^{\text{DL}}+1:kN_a^{\text{DL}})} \mathbf{H}. \end{aligned} \quad (36)$$

Herein, $\mathbf{H}_e^{\text{DL}} \in \mathbb{C}^{N_g^{\text{DL}} \times N_{\text{BS}}}$ is the effective DL channel matrix. Moreover, the RF combiner satisfies $\|\mathbf{A}_{\text{SW}}^{\text{DL}} \mathbf{A}_{\text{PS}}^{\text{DL}}\|_2 = 1$ and thus, the \mathbf{n}'' entries have i.i.d $\mathcal{CN}(0, \sigma^2)$ distribution. We design the precoder \mathbf{P} at the BS to zero force the effective DL channel as

$$\mathbf{P} = \mathbf{H}_e^{\text{DLH}} \left(\mathbf{H}_e^{\text{DL}} \mathbf{H}_e^{\text{DLH}} \right)^{-1}. \quad (37)$$

where \mathbf{P} can be implemented as a fully connected hybrid architecture with no performance penalty, according to [28]. The k^{th} entry of \mathbf{y}^{DL} can be expressed as

$$y_k^{\text{DL}} = \begin{cases} \sqrt{\alpha^{\text{DL}}} \mathbf{P}_t a_H^{\text{DL}} x^{\text{DL}} + n_k'' & \text{if } s_{ik}^{\text{DL}} = 1 \\ \sqrt{\alpha^{\text{DL}}} \mathbf{P}_t a_L^{\text{DL}} x^{\text{DL}} + n_k'' & \text{if } s_{ik}^{\text{DL}} = 0 \end{cases} \quad (38)$$

where n_k'' is the k^{th} entry of \mathbf{n}'' . Note that, the ZF precoder in Eq (37) could increase the received noise power in case of imperfect CSI at the BS as discussed in [32].

A. Downlink detection

For the sake of improving the EE at the UT, we consider energy efficient UT circuitry as depicted in Fig. 1 and thus, avoid applying ML detection. Instead, we propose using the reduced complexity detection method with ERC presented in Section III-C : we exploit the analog devices (AD and 1-bit ADCs) to detect the DL spatial symbol and the digital devices (RF chain and high resolution ADC) to detect the DL modulation symbol.

The k^{th} AD connected to the k^{th} receive analog combiner in Fig. 1 measures the amplitude of the k^{th} signal $|y_k^{\text{DL}}|$ in the RF domain and next, we detect the k^{th} DL spatial bit by comparing the measured amplitude to a threshold through the k^{th} 1-bit ADC

$$\hat{s}_{ik}^{\text{DL}} = \begin{cases} 1 & \text{if } |y_k^{\text{DL}}| \geq \gamma^{\text{DL}} \\ 0 & \text{if } |y_k^{\text{DL}}| < \gamma^{\text{DL}} \end{cases} \text{ with}$$

$$\gamma^{\text{DL}} = \frac{1}{2} \sqrt{\alpha^{\text{DL}}} \mathbf{P}_t (a_H^{\text{DL}} + a_L^{\text{DL}}). \quad (39)$$

Similar to the UL reduced complexity modulation symbol detection, we combine all the signals of \mathbf{y}_k^{DL} . Thereafter, the combined signal passes through the RF chain and the high resolution ADC to detect the DL modulation symbol. The combined signal can be expressed as

$$y_c^{\text{DL,ERC}} = \mathbf{1}_{N_g^{\text{DL}}}^T \mathbf{y}^{\text{DL}} = \sqrt{\alpha^{\text{DL}}} \mathbf{P}_t \mathbf{1}_{N_g^{\text{DL}}}^T \mathbf{t}_t^{\text{DL}} + \mathbf{1}_{N_g^{\text{DL}}}^T \mathbf{n}'' \quad (40)$$

B. Downlink spectral efficiency

The reduced complexity detection method in the DL given with Eq. (39) and Eq. (40) and in the UL given with Eq. (18) and Eq. (21) are similar. Therefore, the SE of the DL transmission $\text{SE}^{\text{DL}} = I_S^{\text{DL}} + I_M^{\text{DL}}$ can be derived in similar way according to Eq. (23) and Eq. (27)

$$I_S^{\text{DL}} = N_g^{\text{DL}} \left(\mathcal{H} \left(\frac{P_0+1-P_1}{2} \right) - \frac{\mathcal{H}(P_0)+\mathcal{H}(1-P_1)}{2} \right), \quad (41)$$

where the sum in Eq. (23) is not needed as the noise power is the same for all groups and

$$\begin{aligned} P_1 &= \Pr(\hat{s}_{ik}^{\text{DL}} = 1 | s_{ik}^{\text{DL}} = 0) \\ &= 1 - Q_1 \left(\sqrt{2\alpha^{\text{DL}}} \mathbf{P}_t a_H^{\text{DL}} / \sigma, \sqrt{2}\gamma^{\text{DL}} / \sigma \right), \\ P_0 &= \Pr(\hat{s}_{ik}^{\text{DL}} = 0 | s_{ik}^{\text{DL}} = 1) \\ &= Q_1 \left(\sqrt{2\alpha^{\text{DL}}} \mathbf{P}_t a_L^{\text{DL}} / \sigma, \sqrt{2}\gamma^{\text{DL}} / \sigma \right), \end{aligned}$$

$$I_M^{\text{DL}} = \frac{1}{2^{N_g^{\text{DL}}}} \sum_{i=1}^{2^{N_g^{\text{DL}}}} \frac{1}{2} \log_2 \left(\frac{4\pi}{e} \text{SNR}_i^{\text{DL}} \right) \text{ with}$$

$$\text{SNR}_i^{\text{DL}} = \frac{|\mathbf{1}_{N_g^{\text{DL}}}^T \mathbf{t}_i^{\text{DL}}|^2}{N_g^{\text{DL}}} \frac{\alpha^{\text{UL}} \mathbf{P}_t}{\sigma^2}. \quad (42)$$

Similar to the UL case, we select the amplitude levels $(a_H^{\text{DL}}, a_L^{\text{DL}})$ to maximize the SE^{DL} .

V. SYSTEM OPTIMIZATION

A. Low complexity uplink/downlink optimization algorithm

In HTSM/HRSM systems, we consider to include an analog phase shifting stage to achieve high gain. On the other side, employing many phase shifters increases the power consumption and could degrade the EE. Thus, we design a hybrid system that reaches the maximum EE such that its SE is equal or larger than the SE achieved by (GTSM/GRSM, same architecture as in Fig. 1 but without phase shifters [36]). In Algorithm 1, we evaluate the SE of the GTSM/GRSM systems for comparison with the proposed hybrid systems and as an input to Algorithm 2. We apply the QR decomposition [40] to sort the channel matrix rows. Specifically, the set \mathcal{A} in step 3 of Algorithm 1 includes the most uncorrelated UT antennas

Algorithm 1 UL and DL system parameters optimization of generalized SM

- 1: **Input** : \mathbf{H} and $N_{g,\max}$
 - 2: **Output** : $\text{SE}_{\text{GTSM}}^{\text{UL}}$ and $\text{SE}_{\text{GRSM}}^{\text{DL}}$
 - 3: $[\mathbf{Q} \ \mathbf{R} \ \mathcal{A}] = \text{QR}(\mathbf{H}^H, 0)$ such that $\mathbf{H}^H(:, \mathcal{A}) = \mathbf{QR}$
 - 4: **for** $N_g = 1 : N_{g,\max}$
 - 5: $\mathbf{H}_e^{\text{DL}} = [\mathbf{H}^{(\mathcal{A}(1),:)}; \dots; \mathbf{H}^{(\mathcal{A}(N_g),:)}], \mathbf{H}_e^{\text{UL}} = \mathbf{H}_e^{\text{DLH}}$
 - 6: $\mathcal{SE}_{\text{GTSM}}^{\text{UL}}(N_g) = \underset{a_0^{\text{UL}}}{\text{maximize}} (I_S^{\text{UL}} + I_M^{\text{UL}}), \text{ s.t. } 0 \leq a_0^{\text{UL}} \leq \frac{1}{2}$
 - 7: $\mathcal{SE}_{\text{GRSM}}^{\text{DL}}(N_g) = \underset{a_0^{\text{DL}}}{\text{maximize}} (I_S^{\text{DL}} + I_M^{\text{DL}}), \text{ s.t. } 0 \leq a_0^{\text{DL}} \leq \frac{1}{2}$
 - 8: **end for**
 - 9: **return** $\text{SE}_{\text{GTSM}}^{\text{UL}} = \max(\mathcal{SE}_{\text{GTSM}}^{\text{UL}})$
 $\text{SE}_{\text{GRSM}}^{\text{DL}} = \max(\mathcal{SE}_{\text{GRSM}}^{\text{DL}})$
-

Algorithm 2 UL and DL system parameters optimization of the hybrid SM

- 1: **Input** : $\mathbf{H} = \mathbf{A}_R \mathbf{D} \mathbf{A}_i^H$, $\text{SE}_{\text{GTSM}}^{\text{UL}}$, $\text{SE}_{\text{GRSM}}^{\text{DL}}$ and P_t
 - 2: **Output** : $\mathbf{A}_{\text{SW}}^{\text{UL}*}$, $\mathbf{A}_{\text{PS}}^{\text{UL}*}$, $\mathbf{A}_{\text{SW}}^{\text{DL}*}$, $\mathbf{A}_{\text{PS}}^{\text{DL}*}$, $\alpha_0^{\text{UL}*}$ and $\alpha_0^{\text{DL}*}$.
 - 3: $\theta_{\max} = \theta_i : i = \arg \max_j |\mathbf{D}_{j,j}|, j = 1, \dots, L$
 - 4: **for** $N_a = 1 : N_U$
 - 5: Generate all possible antennas arrays sub-channels $\mathbf{H}_i = \mathbf{A}_{\text{SW},i} \mathbf{H}, i = 1, \dots, K = \binom{N_U}{N_a}, \mathbf{A}_{\text{SW},i}^{(l,m)} \in \{0, 1\}, \|\mathbf{A}_{\text{SW},i}^{(k)}\|_0 = 1, k = 1, \dots, N_a, \|\mathbf{A}_{\text{SW},i}^{(i,:)}\|_0 \in \{0, 1\}, i = 1, \dots, N_U$.
 - 6: $\mathbf{f}_i = \sqrt{\frac{N_r}{N_a}} \mathbf{A}_{R,i}(:, \theta_{\max}), i = 1, \dots, K$
 - 7: $\mathbf{H}_e = [\mathbf{f}_1^H \mathbf{H}_1; \dots; \mathbf{f}_K^H \mathbf{H}_K]$
 - 8: $[\mathbf{Q} \ \mathbf{R} \ \mathcal{A}] = \text{QR}(\mathbf{H}_e^H, 0)$ such that $\mathbf{H}_e^H(:, \mathcal{A}) = \mathbf{QR}$
 - 9: **for** $N_g = 1 : N_{g,\max}$
 - 10: $\mathbf{A}_{\text{SW}}(N_a, N_g) = [\mathbf{A}_{\text{SW},\mathcal{A}(1)}, \dots, \mathbf{A}_{\text{SW},\mathcal{A}(N_g)}]$
 - 11: $\mathbf{A}_{\text{PS}}(N_a, N_g) = \text{blockdiag}(\mathbf{f}_{\mathcal{A}(1)}, \dots, \mathbf{f}_{\mathcal{A}(N_g)})$
 - 12: $\mathbf{H}_e^{\text{UL}} = [\mathbf{H}_{\mathcal{A}(1)}^H \mathbf{f}_{\mathcal{A}(1)}, \dots, \mathbf{H}_{\mathcal{A}(N_g)}^H \mathbf{f}_{\mathcal{A}(N_g)}]$
 - 13: $\mathcal{SE}^{\text{UL}}(N_a, N_g) = \underset{a_0^{\text{UL}}}{\text{maximize}} (I_S^{\text{UL}} + I_M^{\text{UL}}) \text{ s.t. } 0 \leq a_0^{\text{UL}} \leq \frac{1}{2}$
 - 14: $\Omega^{\text{UL}}(N_a, N_g) = a_0^{\text{UL}}$
 - 15: $\mathbf{H}_e^{\text{DL}} = [\mathbf{f}_{\mathcal{A}(1)}^H \mathbf{H}_{\mathcal{A}(1)}; \dots; \mathbf{f}_{\mathcal{A}(N_g)}^H \mathbf{H}_{\mathcal{A}(N_g)}]$
 - 16: $\mathcal{SE}^{\text{DL}}(N_a, N_g) = \underset{a_0^{\text{DL}}}{\text{maximize}} (I_S^{\text{DL}} + I_M^{\text{DL}}) \text{ s.t. } 0 \leq a_0^{\text{DL}} \leq \frac{1}{2}$
 - 17: $\Omega^{\text{DL}}(N_a, N_g) = a_0^{\text{DL}}$
 - 18: **end for**
 - 19: **end for**
 - 20: Solve the optimization problem in Eq. 43 to obtain the operating points (N_a^*, N_g^*) for the UL and the DL transmissions
 - 21: **return** $\mathbf{A}_{\text{SW}}^{\text{UL}*} = \mathbf{A}_{\text{SW}}(N_a^{\text{UL}*}, N_g^{\text{UL}*}), \mathbf{A}_{\text{PS}}^{\text{UL}*} = \mathbf{A}_{\text{PS}}(N_a^{\text{UL}*}, N_g^{\text{UL}*}),$
 $\alpha_0^{\text{UL}*} = \Omega(N_a^{\text{UL}*}, N_g^{\text{UL}*}), \mathbf{A}_{\text{SW}}^{\text{DL}*} = \mathbf{A}_{\text{SW}}(N_a^{\text{DL}*}, N_g^{\text{DL}*}), \mathbf{A}_{\text{PS}}^{\text{DL}*} =$
 $\mathbf{A}_{\text{PS}}(N_a^{\text{DL}*}, N_g^{\text{DL}*}), \alpha_0^{\text{DL}*} = \Omega(N_a^{\text{DL}*}, N_g^{\text{DL}*}).$
-

sorted in descending order according to the strength of the path between the UT antenna and the BS and $N_{g,\max}$ represents the maximum number of groups, a value that is upper bounded by the number of channel clusters C . Next, we select the best sub-channel matrix that maximizes the SE^2 . In Algorithm 2, we optimize the analog beamforming and combining matrices and the spatial amplitude levels in UL and DL at a given P_t . Specifically, for a given number of phase shifters (N_a) inside the group, we have $\binom{N_U}{N_a}$ possible ways of connecting the N_a phase shifters to the N_U antennas of the UT and this leads to

²We cannot consider the full channel matrix when the matrix is rank deficient as the ZF precoding/combining does not exist.

$\binom{N_U}{N_a}$ possible different groups of phase shifters. Each phase shifters group is designed to steer the beam in the direction of the strongest path. Next, we generate a large effective channel matrix that includes all of the possible antenna arrays groups. Thanks to the QR decomposition [40], we can sort the linearly independent groups in one step³. Note that in Algorithm 2, steps 5 through 8 are common for the UL and DL. Thereafter, we evaluate the UL and DL SE and the EE (defined as the SE divided by the UT hardware power consumption) with optimized amplitude levels for number of groups starts from one to $N_{g,\max}$. We repeat the procedure for every number of active antennas in a group ($N_a = 1 : N_U$) until we complete $N_U \times N_{g,\max}$ grids for the SE^{UL} , SE^{DL} , EE^{UL} and EE^{DL} . Finally, the BS selects the UL and DL operating points in the grids that maximize the EE such that the SE is better than that of systems without phase shifters (GTSM/GRSM) evaluated in Algorithm 1. This is formulated as:

$$(N_a^*, N_g^*) = \underset{N_a \in \{1, \dots, N_U\}, N_g \in \{1, \dots, N_{g,\max}\}}{\text{maximize}} \text{EE} = \frac{\text{SE}(N_a, N_g, P_t)}{P_C(N_a, N_g, P_t)}$$

$$\text{subject to} \quad \text{SE} \geq t \quad (43)$$

where problem (43) is solved for the UL considering $\text{SE} = \mathcal{SE}^{\text{UL}}, P_C = P_C^{\text{UL}}, t = \text{SE}_{\text{GTSM}}^{\text{UL}}$ and for the DL considering $\text{SE} = \mathcal{SE}^{\text{DL}}, P_C = P_C^{\text{DL}}$ and $t = \text{SE}_{\text{GRSM}}^{\text{DL}}$. The optimization of the amplitude levels in steps 13 and 17 of Algorithm 2 leads to non linear objective function in one unknown and one linear constraint that can be efficiently evaluated using bisection method. Algorithm 2 maximizes the EE at a given transmit power, but the optimal transmit power still needs to be computed. This can be done in two steps. First, evaluating the the minimum transmit power that ensures the SE constraint of problem (43). Second, determine the transmit power within the evaluated feasible interval that maximizes the EE. In step 1, we apply the bisection method in Algorithm 3 with initial lower bound ($P_t = 0$) and upper bound ($P_t = \text{maximum transmit power } P_{t,\max}$). At each iteration, we apply Algorithm 2 using the value of P_t in the middle of the upper and lower bounds. The updated lower bound is P_t if problem (43) is infeasible, otherwise, the updated upper bound is P_t . The bisection iterations stop when the gap between the bounds satisfies specific accuracy. The output of step 1 is the minimum transmit power $P_{t,\min}$ that ensures the SE constraint of problem (43). In step 2, we apply another bisection method as illustrated in Algorithm 3 with initial lower bound ($P_t = 0$) and upper bound ($P_t = P_{t,\min}$). At each iteration, we solve Algorithm 2 at P_t in the middle of the bounds. The updated lower bound is P_t if the optimized EE at P_t is greater than or equals to the optimized EE at the lower bound, otherwise, the updated upper bound is P_t . The iterations stop when the gap between the bounds satisfies a given accuracy level. The output of step 2 is the optimal transmit power that maximizes the EE under SE constraint. In exhaustive search based design, line 8 in Algorithm 2 should

³Note that we need linearly independent groups to perform ZF precoding/combining matrices.

Algorithm 3 Transmit power optimization using bisection method

- 1: **Input** : lower bound, upper bound and ϵ
- 2: **if** $|\text{lower bound} - \text{upper bound}| > \epsilon$ **then**
- 3: Apply Algorithm 2 at $P_t = \frac{\text{lower bound} + \text{upper bound}}{2}$
- 4: **if** optimization condition is satisfied, problem (43) is infeasible for step 1 or the EE resulting from solving problem (43) at $P_t = \frac{\text{lower bound} + \text{upper bound}}{2}$ is greater than or equals to the EE at $P_t = \text{lower bound}$ for step 2 **then**
- 5: lower bound = P_t , else, upper bound = P_t
- 6: **end if**
- 7: **end if**

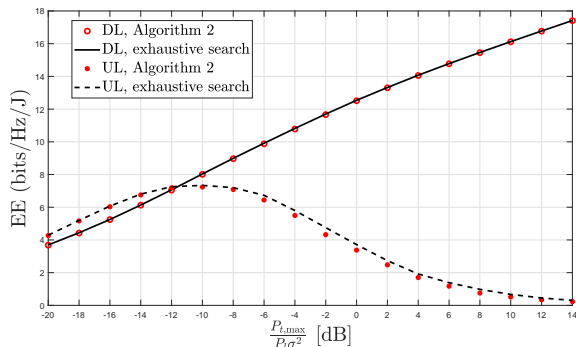


Fig. 3: EE of the proposed schemes evaluated by Algorithm 2 and by exhaustive search vs. received SNR at $N_{BS} = 128$, $N_U = 4$, $N_{g,max} = C = 3$ (average over 1000 channel realizations).

be replaced with exhaustive search selection of N_g groups out of $\binom{N_u}{N_a}$ possible groups for all values of N_g . The number of grid points of the proposed algorithm $N_{g,p}$ and the exhaustive search $N_{g,es}$ can be expressed as

$$N_{g,p} = N_U \times N_{g,max},$$

$$N_{g,es} = 1 + \sum_{N_a=1}^{N_U-1} \sum_{N_g=1}^{N_{g,max}} \binom{N_U}{N_a}, \forall \binom{N_U}{N_a} \geq N_g. \quad (44)$$

As an illustrative example, consider $N_U = 16$ and $N_{g,max} = C = 4$. The corresponding number of grid points are $N_{g,p} = 64$ and $N_{g,es} = 2.9 \times 10^{15}$ and thus, the proposed algorithm significantly reduces the computational complexity.

VI. SIMULATION RESULTS

In this section, we evaluate the performance of the proposed HTSM/HRSM schemes compared to the GTSM/GRSM schemes in terms of SE and EE. We show the achievable SE, EE, SE-EE trade-off and illustrate the optimal numbers of needed groups and phase shifters per group for the UL and the DL systems. In order to validate the efficiency of the proposed algorithm, we compare the EE obtained from the reduced complexity algorithm with the one obtained from the exhaustive search. We evaluate the system performance in both stochastic and deterministic channel environments.

A. Performance evaluation in stochastic channel

In the stochastic simulation environment, we consider $\sigma^2 = -84$ dBm and $P_t = 90$ dB.

Fig. 3 shows the EE comparison of the UL HTSM and the DL HRSM schemes when we apply the proposed fast

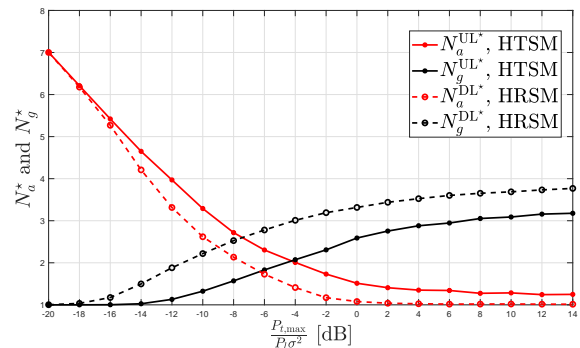


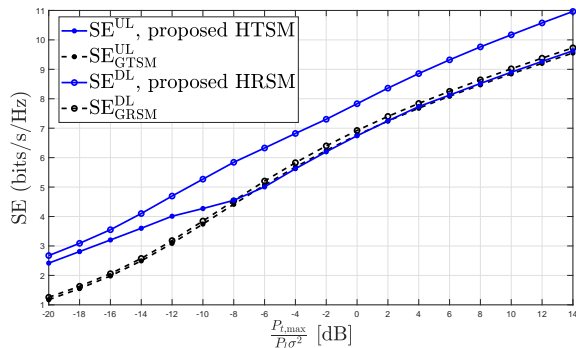
Fig. 4: Optimum number of antennas per group and number of groups of the HTSM/HRSM schemes vs. received SNR at $N_{BS} = 128$, $N_U = 16$, $N_{g,max} = C = 4$ (average over 1000 channel realizations).

Algorithm 2 compared to the exhaustive search. Thanks to the QR decomposition in Algorithm 2, we obtain the same performance as the exhaustive search with significant reduction in the computational complexity as explained in Eq. (44).

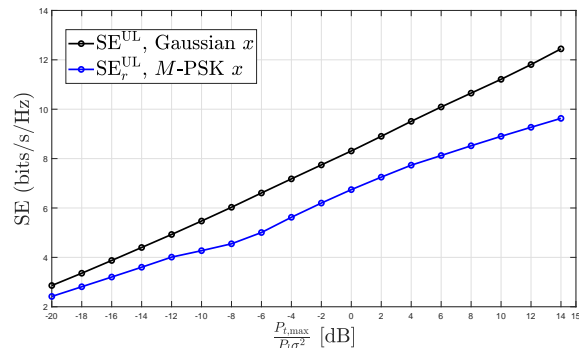
Fig. 4 shows the proposed system behavior in terms of the optimized number of groups and antennas per groups of the HTSM/HRSM designs. The number of groups and antennas per group are obtained from Algorithm 2 and are designed to ensure full rank effective channel matrix and enable the ZF combining and precoding in the UL and DL, respectively. As we maximize the EE, we keep the total number of the phase shifters small. Therefore, the increase in number of groups is necessarily associated to the decrease in the number of phase shifters per group. At low SNR, we need high beamforming/combining gains. Hence, the number of phase shifters per group is high and thus, the number of groups is small. Increasing the SNR reduces the required beamforming/combining gains. As a result, the number of phase shifters per group decreases and the number of groups increases to attain high spatial multiplexing gain.

Fig. 5a shows the SE of the proposed UL HTSM and DL HRSM designs compared to the UL GTSM and the DL GRSM schemes. At low SNR regime (common assumption associated with outdoor mmWave propagation), the proposed hybrid designs achieve superior SE as the phase shifters stages in the HTSM and HRSM schemes provide high beamforming and combining gains; respectively and combat the severe path-loss. At high SNR, each group may contain one or two phase shifters as explained in Fig. 4. Since the small number of antennas at an array is not sufficient to provide high beamforming gains, the GTSM approaches the SE of the HTSM. On the other hand, the HRSM still outperforms GRSM at high SNR even with the small number of phase shifters per group. The SE of the HRSM is higher than HTSM as the ZF combiner in the HTSM system could amplify the UL noise power. In contrast, the RF combiner of the HRSM does not affect the DL noise power.

Fig. 5b shows the SE of the UL HTSM scheme when we apply Algorithm 2 with (Gaussian input distribution and optimum detector) shown in Eq. (11) and Eq. (17) and (M -PSK modulated input and reduced complexity detector) as in Eq. (23) and Eq. (27). The reduced complexity scheme approaches the optimal performance specifically at low SNR.

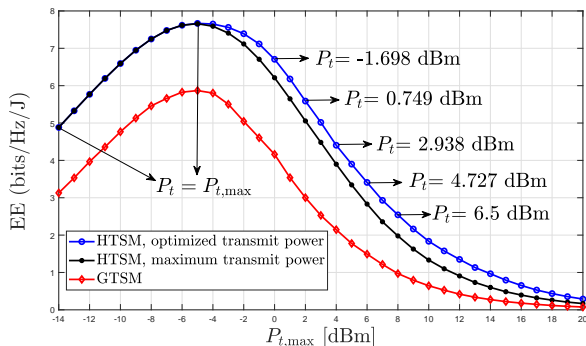


(a) HTSM-HRSM compared to GTSM-GRSM

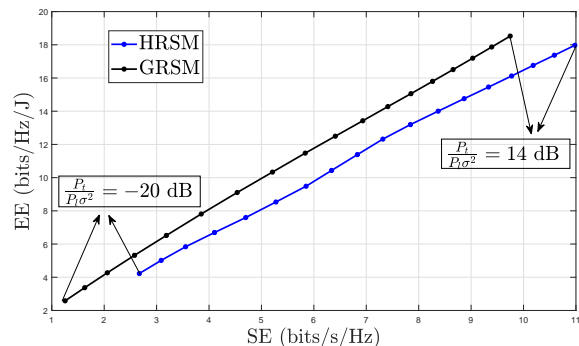


(b) Optimal and reduced complexity detectors

Fig. 5: SE of the proposed HTSM-HRSM compared to GTSM-GRSM schemes and SE of HTSM with Gaussian modulation symbol and optimal detector compared to the scheme with M -PSK modulation symbol and reduced complexity detector at $N_{BS} = 128$, $N_U = 16$, $N_{g,max} = C = 4$ (average over 1000 channel realizations).

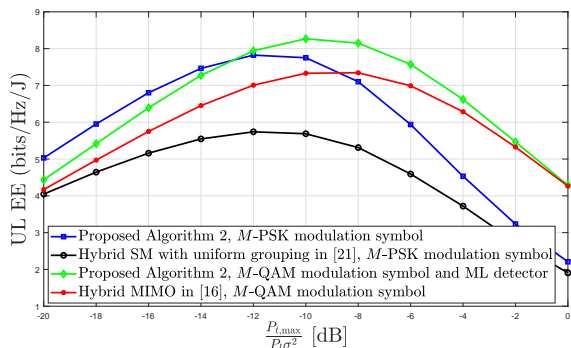


(a) UL EE

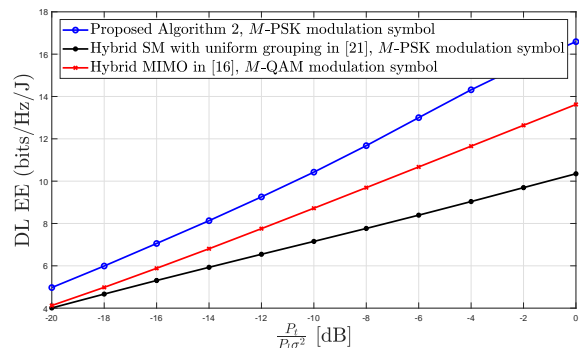


(b) DL EE-SE trade-off

Fig. 6: UL EE and DL EE-SE trade-off of the proposed HTSM and HRSM schemes compared to GTSM and GRSM methods at $N_{BS} = 128$, $N_U = 16$, $N_{g,max} = C = 4$ (average over 1000 channel realizations).



(a) UL EE



(b) DL EE

Fig. 7: UL and DL energy efficiency at the UT of the proposed scheme compared to hybrid SM in [21] and hybrid MIMO in [16] assuming single RF chain, $N_{BS} = 128$, $N_U = 16$, $N_{g,max} = C = 6$ (average over 1000 channel realizations).

Fig. 6a shows the UL EE of the proposed HTSM scheme when the transmit power is optimized, or the maximum available transmit power is used. We compare it to the UL EE of the GTSM system. The proposed scheme outperforms the GTSM system especially at low SNR when the beamforming gain is needed. Optimizing the transmit power slightly improves the EE due to the SE constraint shown in Eq. (43). Fig. 6b represents the DL EE-SE trade-off of the proposed HRSM design compared to the GRSM scheme. At low SNR, the proposed hybrid design achieves superior SE and EE due to the high gain of the phase shifters stage. At high SNR regime, generalized system achieves slightly higher EE as the

number of groups tends to be one and the SE gap of the two systems reduces.

Fig. 7 shows the UL and DL EE at the UT of the proposed scheme compared to hybrid SM in [21] and hybrid MIMO in [16] assuming single RF chain at the UT. Transmitting M -PSK modulation symbols and applying reduced complexity detector, the proposed hybrid SM with optimized grouping explained in Algorithm 2 attains higher EE than hybrid SM with uniform grouping proposed in [21] in UL and DL. Moreover, in DL, it outperforms the the hybrid MIMO in [16]. Considering M -QAM symbols and applying ML detector, the proposed UL HTSM achieves superior EE than the hybrid

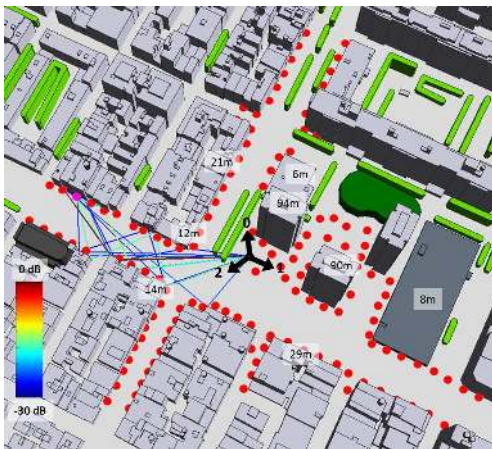


Fig. 8: Top view of realistic users distribution (red dots) served by three sectors mMIMO BSs inside mmWave small-cell in Manhattan area in New York City where the farthest user at 220 metre distance from the small-cell.

MIMO in [16].

B. Performance evaluation for a ray-trace model

With the aim of evaluating the performance of the proposed system in typical small-cell scenario at 28 GHz, we consider a realistic user distribution and generate deterministic channels per user, and then compare the system performance with the stochastic and deterministic channels. Several outdoor small-cell mmWave channel samples have been predicted from the ray-based propagation model VolcanoUrban [33]. Those samples are the result of physical interactions between the electromagnetic wave and the real representation of a dense urban environment, more precisely, a district in New York Manhattan. A small-cell is positioned at 8 meters above the ground, at a typical location for a lamppost. Three sectors are installed at the small-cell. Each sector is feeding a linear antenna array with boresight direction oriented towards azimuth 0° , 120° and 240° , as depicted in Fig. 8. Each linear antenna array is formed of 128° vertically-polarized antenna elements, which are uniformly distributed in the horizontal plane, at frequency 28 GHz, and with half-wavelength separation. All antenna elements have same radiation pattern with 60° half-power horizontal beamwidth.

The users are assumed to be pedestrians distributed on the surrounding pavements at a maximum 220 meters range from the small-cell. The user equipment is located at 1.5 meter above the ground. Its antenna is a uniform linear array with 16 vertically-polarized isotropic elements positioned in the horizontal plane. The channel samples are produced from 142 different user positions. Users are positioned either in a wide or a narrow street, or even in a small square. Few of them are in non line-of-sight (NLoS) situation. Finally, a total number of 180 channel samples are created: 50, 69 and 61 for respectively

sector 0, 1 and 2 with 37 NLoS samples.

The SE of the proposed UL HTSM and DL HRSM designs evaluated on the stochastic channel model with $C = 2$ and $C = 6$ scatterers and the deterministic channel model for the scenario proposed in Fig. 8 assuming the same path-loss for the two models is depicted in Fig. 9. The noise level is $\sigma^2 = -84$ dBm, the transmit power is $P_t = 20$ dBm, the carrier frequency is $f_c = 28$ GHz, bandwidth $BW = 10$ MHz and 76% of the users have delay spread smaller than the symbol time for the simulation setup so that we can consider non-frequency selective channel. Sector 0 has the lowest scattering environment due to the LoS users and the vegetation. Sector 1 has more NLoS users and thus, the users achieve high SE. Sector 2 users are farther away than the users in the other sectors and thus, its users have greater path-loss and lower SE. From this experiment, we show that the proposed design not only attains high performance with the theoretical stochastic channel model in Eq. (3) but it also achieves similar performance with the realistic channel model. Moreover, the performance evaluation of the 28 GHz channels at $BW = 10$ MHz based on stochastic channel model gives a realistic assessment if the number of clusters is in the range of $C = 2$ and $C = 6$.

VII. CONCLUSION

In this paper, we proposed novel and energy efficient hybrid transceiver architecture based on two stages analog beamformer in the UL and combiner in the DL, respectively. The analog switches stage smartly allocate the UT antennas on the phase shifters groups to minimize the spatial correlation. Moreover, the analog phase shifters stage maximizes the beamforming/combining gains to combat the path-loss. We proposed a novel and computationally efficient optimization algorithm to design the analog stages. The proposed design achieves the same performance as the exhaustive search method but with much lower computational complexity. The flexibility of the architecture allows optimising the hybrid transceiver at any SNR regime: At low SNR regime, we activate only one group of phase shifters and maximize the number of phase shifters inside the group to attain high post processing SNR. At high SNR regime, the number of groups increases and as a result the spatial rate increases. Moreover, the number of phase shifters per group decreases as optimizing the EE implies reducing the total number of phase shifters. We validated the performance of the proposed design on a realistic deployment in Manhattan area in New York City. The performance evaluation for mmWave small-cell at 28 GHz shows that the stochastic channel models provides results close to those obtained with the deterministic channel if the number of clusters is chosen to emulate the real-world scenario.

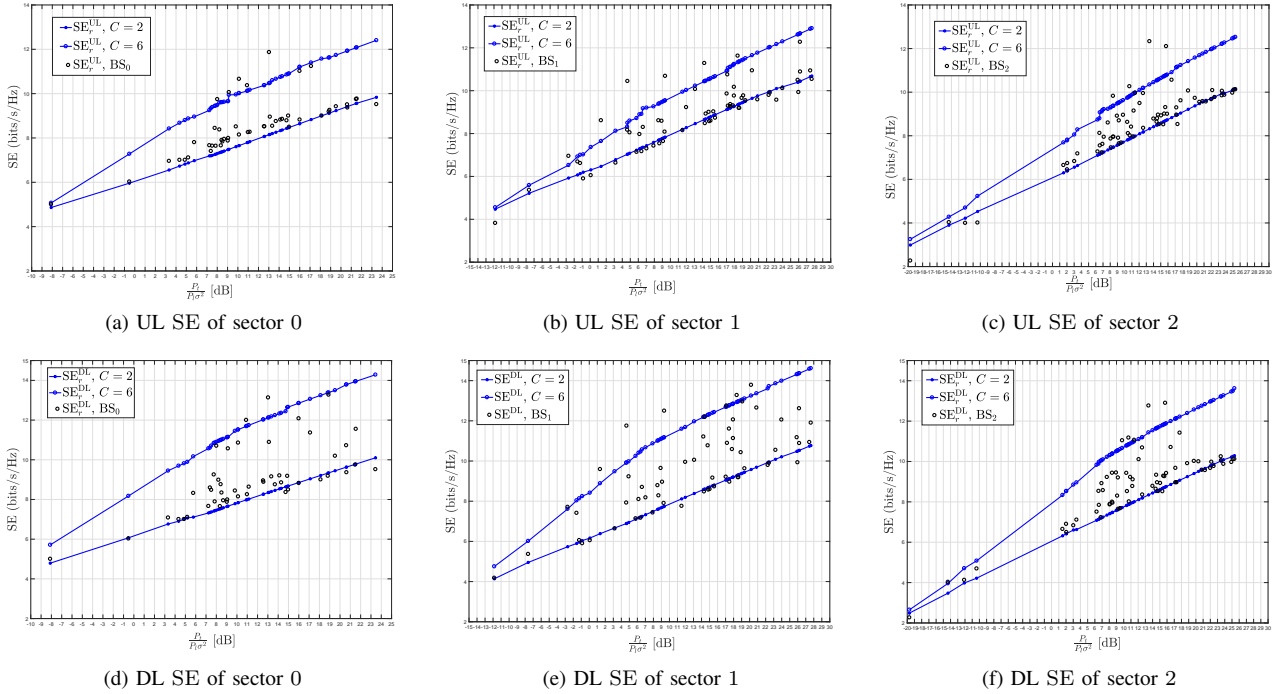


Fig. 9: SE of the proposed UL and DL hybrid design evaluated on stochastic (in blue, for $C = 2$, $C = 6$ and average over 100 realizations) and deterministic channel samples (in dots) assuming the same path-loss for the two models, $\sigma^2 = -84$ dBm, $P_t = 20$ dBm, $f_c = 28$ GHz and BW = 10 MHz.

APPENDIX A

Proof of Lemma 1

$$\begin{aligned}
& h\left(y_k^{\text{UL}} | y_{k-1}^{\text{UL}}, \dots, y_1^{\text{UL}}, \mathbf{t}_i^{\text{UL}(1:k)}\right) \stackrel{(a)}{=} h\left(y_1^{\text{UL}}, \dots, y_k^{\text{UL}} | \mathbf{t}_i^{\text{UL}(1:k)}\right) - h\left(y_1^{\text{UL}}, \dots, y_{k-1}^{\text{UL}} | \mathbf{t}_i^{\text{UL}(1:k)}\right) \\
& \stackrel{(b)}{=} \log_2 \pi e \left| \alpha^{\text{UL}} P_t \mathbf{t}_i^{\text{UL}(1:k)} \mathbf{t}_i^{\text{UL}(1:k)H} + \mathbf{R}_{n'n'}^{(1:k,1:k)} \right| \left| \alpha^{\text{UL}} P_t \mathbf{t}_i^{\text{UL}(1:k-1)} \mathbf{t}_i^{\text{UL}(1:k-1)H} + \mathbf{R}_{n'n'}^{(1:k-1,1:k-1)} \right| \\
& \stackrel{(c)}{=} \log_2 \pi e \frac{\left| \mathbf{R}_{n'n'}^{(1:k,1:k)} \right|}{\left| \mathbf{R}_{n'n'}^{(1:k-1,1:k-1)} \right|} \frac{1 + \alpha^{\text{UL}} P_t \mathbf{t}_i^{\text{UL}(1:k)T} \mathbf{R}_{n'n'}^{(1:k,1:k)^{-1}} \mathbf{t}_i^{\text{UL}(1:k)}}{1 + \alpha^{\text{UL}} P_t \mathbf{t}_i^{\text{UL}(1:k-1)T} \mathbf{R}_{n'n'}^{(1:k-1,1:k-1)^{-1}} \mathbf{t}_i^{\text{UL}(1:k-1)}} \\
& \stackrel{(d)}{=} \log_2 \pi e \left(\sigma_K^2 + P_{i,K} \right), \text{ where } \sigma_K^2 = \left| \mathbf{R}_{n'n'}^{(1:k,1:k)} \right| \left| \mathbf{R}_{n'n'}^{(1:k-1,1:k-1)} \right| \text{ and} \\
& P_{i,K} = \sigma_K^2 \alpha^{\text{UL}} P_t \frac{\mathbf{t}_i^{\text{UL}(1:k)T} \mathbf{R}_{n'n'}^{(1:k,1:k)^{-1}} \mathbf{t}_i^{\text{UL}(1:k)} - \mathbf{t}_i^{\text{UL}(1:k-1)T} \mathbf{R}_{n'n'}^{(1:k-1,1:k-1)^{-1}} \mathbf{t}_i^{\text{UL}(1:k-1)}}{1 + \alpha^{\text{UL}} P_t \mathbf{t}_i^{\text{UL}(1:k-1)T} \mathbf{R}_{n'n'}^{(1:k-1,1:k-1)^{-1}} \mathbf{t}_i^{\text{UL}(1:k-1)}}
\end{aligned}$$

Step $\stackrel{(a)}{=}$ follows from chain rule of entropy, step $\stackrel{(b)}{=}$ follows from Gaussian distributions, step $\stackrel{(c)}{=}$ follows from applying the identity $|\mathbf{A} + \mathbf{t}\mathbf{t}^H| = |\mathbf{A}| (1 + \mathbf{t}^H \mathbf{A}^{-1} \mathbf{t})$ and step $\stackrel{(d)}{=}$ follows from noise and signal powers separation. Therefore,

$$f\left(y_k^{\text{UL}} | y_{k-1}^{\text{UL}}, \dots, y_1^{\text{UL}}, \mathbf{t}_i^{\text{UL}(1:k)}\right) = \mathcal{CN}\left(0, \sigma_K^2 + P_{i,K}\right).$$

REFERENCES

- [1] *International Telecommunication Union (ITU). IMT traffic estimates for the years 2020 to 2030*, Report ITU-R M.2370-0, Jul. 2015.
- [2] T. S. Rappaport, R. W. Heath Jr., R. C. Daniels, and J. N. Murdock, *Millimeter wave wireless communications*. Englewood Cliffs, NJ, USA: Prentice-Hall, Sep. 2014.
- [3] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 101–107, Jun. 2011.
- [4] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proceedings of the IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.
- [5] T. L. Marzetta, "Massive MIMO: An introduction," *Bell Labs Technical Journal*, vol. 20, pp. 11–22, Mar. 2015.
- [6] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, Jan. 2013.
- [7] L. Sanguinetti, E. Björnson, M. Debbah, and A. L. Moustakas, "Optimal linear precoding in multi-user MIMO systems: A large system analysis," in *IEEE Global Communications Conference (GLOBECOM)*, pp. 3922–3927, Dec. 2014.
- [8] N. Shariati, E. Björnson, M. Bengtsson, and M. Debbah, "Low-complexity channel estimation in large-scale MIMO using polynomial expansion," in *24th IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pp. 1157–1162, Sep. 2013.
- [9] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, Oct. 2014.
- [10] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 461–471, Feb. 2004.
- [11] 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, Apr. 2016.
- [12] O. E. Ayach *et al.*, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.
- [13] A. Alkhateeb, G. Leus, and R. W. Heath Jr., "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Transactions on Wireless Communications*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.
- [14] 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.
- [15] F. Söhrabi 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.
- [16] X. Yu *et al.*, "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.
- [17] R. Y. Mesleh *et al.*, "Spatial modulation," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 4, pp. 2228–2241, Jul. 2008.
- [18] R. Zhang, L.-L. Yang, and L. Hanzo, "Generalised pre-coding aided spatial modulation," *IEEE Transactions on Wireless Communications*, vol. 12, no. 11, pp. 5434–5443, Nov. 2013.
- [19] M. D. Renzo *et al.*, "Spatial modulation for generalized MIMO: Challenges, opportunities and implementation," *Proceedings of the IEEE*, vol. 102, no. 1, pp. 56–103, Jan. 2014.
- [20] Y. Cui, X. Fang, and L. Yan, "Hybrid spatial modulation beamforming for mmwave railway communication systems," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 12, pp. 9597–9606, Dec. 2016.
- [21] M. Yüzgeçcioğlu and E. Jorswieck, "Hybrid beamforming with spatial modulation in multi-user massive MIMO mmwave networks," in *28th IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pp. 1–6, Oct. 2017.
- [22] 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, Jun. 2014.
- [23] 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.
- [24] A. Raafat *et al.*, "Energy efficient transmit-receive spatial modulation for uplink-downlink large-scale MIMO systems," in the *proceedings of IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, Dec. 2018.
- [25] L. He, J. Wang, and J. Song, "On generalized spatial modulation aided millimeter wave MIMO: Spectral efficiency analysis and hybrid precoder design," *IEEE Trans. on Wireless Communications*, vol. 16, no. 11, pp. 7658–7671, Nov. 2017.
- [26] —, "Spatial modulation for more spatial multiplexing: Rf-chain-limited generalized spatial modulation aided MM-Wave MIMO with hybrid precoding," *IEEE Trans. on Communications*, vol. 66, no. 3, pp. 986–998, Mar. 2018.
- [27] 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.
- [28] O. El Ayach *et al.*, "Low complexity precoding for large millimeter wave MIMO systems," in *IEEE international conference on communications (ICC)*, pp. 3724–3729, Jun. 2012.
- [29] S. Shakib *et al.*, "A highly efficient and linear power amplifier for 28-GHz 5G phased array radios in 28-nm CMOS," *IEEE Journal of Solid-State Circuit*, vol. 51, no. 12, pp. 3020–3036, Dec. 2016.
- [30] J. Lagos *et al.*, "A single-channel, 600-MS/s, 12-b, ringamp-based pipelined ADC in 28-nm cmos," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 2, pp. 403–416, Feb. 2018.
- [31] N. Rostomyan, M. Özen, and P. Asbeck, "28 GHz dohertry power amplifier in cmos soi with 28% back-off pae," *IEEE Microwave and Wireless Components Letters*, vol. 28, no. 5, pp. 446–448, May 2018.
- [32] C. Wang *et al.*, "On the performance of the MIMO zero-forcing receiver in the presence of channel estimation error," *IEEE Transactions on Wireless Communications*, vol. 6, no. 3, pp. 805–810, Mar. 2007.
- [33] Y. Corre and Y. Lostanlen, "Three-dimensional urban EM wave propagation model for radio network planning and optimization over large areas," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 7, pp. 3112–3123, Sep. 2009.
- [34] A. Raafat, A. Agustin, and J. Vidal, "MMSE precoding for receive spatial modulation in large MIMO systems," in *IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 1–5, Jun. 2018.
- [35] M. F. Huber, T. Bailey, H. Durrant-Whyte, and U. D. Hanebeck, "On entropy approximation for Gaussian mixture random vectors," in *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pp. 181–188, Aug. 2008.
- [36] A. Raafat, A. Agustin, and J. Vidal, "Receive spatial modulation for massive MIMO systems," in *IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, Dec. 2017.
- [37] F. Chapeau-Blondeau, "Noise-enhanced capacity via stochastic resonance in an asymmetric binary channel," *Physical Review E*, vol. 55, no. 2, p. 2016, Feb. 1997.
- [38] P. E. McIllree, *Channel capacity calculations for M-ary N-dimensional signal sets*. M.S. thesis, The U. South Australia, School of Electronic Eng., Feb. 1995.
- [39] B. Goebel *et al.*, "Calculation of mutual information for partially coherent Gaussian channels with applications to fiber optics," *IEEE Transactions on Information Theory*, vol. 57, no. 9, pp. 5720–5736, Sep. 2011.
- [40] Y. Wu *et al.*, "Receive antenna selection in the downlink of multiuser MIMO systems," *62nd Vehicular Technology Conference*, vol. 1, pp. 477–481, Sep. 2005.