Radio frequency-chain selection for energy and spectral efficiency maximization in hybrid beamforming under hardware imperfections

Citation for published version:
Vlachos, E, Thompson, J, Kaushik, A & Masouros, C 2020, 'Radio frequency-chain selection for energy and spectral efficiency maximization in hybrid beamforming under hardware imperfections', *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 476, no. 2244, 20200451. https://doi.org/10.1098/rspa.2020.0451

Digital Object Identifier (DOI):
10.1098/rspa.2020.0451

Published In:
Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences

PROCEEDINGS A

rspa.royalsocietypublishing.org

Research

Article submitted to journal

Subject Areas:
electrical engineering

Keywords:
millimetre wave hybrid MIMO, energy efficiency maximisation, RF–chain selection

Author for correspondence:
John Thompson
e-mail: j.s.thompson@ed.ac.uk

RF–Chain Selection for Energy and Spectral Efficiency Maximisation in Hybrid Beamforming under Hardware Imperfections

Evangelos Vlachos[1], John Thompson[2], Aryan Kaushik[3], Christos Masouros[3]

[1]Industrial Systems Institute, ATHENA Research Centre, Greece
[2]Institute for Digital Communications, The University of Edinburgh, UK
[3]Department of Electronic and Electrical Engineering, University College London, UK

(Invited Paper)

The next generation wireless communications requires reduced energy consumption, increased data rates and better signal coverage. The millimetre wave frequency spectrum above 30 GHz can help fulfil the performance requirements of the next generation mobile broadband systems. Multiple-input multiple-output (MIMO) technology can provide performance gains to help mitigate the increased path loss experienced at mmWave frequencies compared to microwave bands. Emerging hybrid beamforming architectures can reduce the energy consumption and hardware complexity with the use of fewer Radio-Frequency (RF) chains. Energy efficiency is identified as a key fifth generation (5G) metric and will have a major impact on the hybrid beamforming system design. In terms of transceiver power consumption, deactivating parts of the beamformer structure to reduce power typically leads to significant loss of spectral efficiency. Our aim is to achieve the highest energy efficiency for the millimetre-wave communications system while mitigating the resulting loss in spectral efficiency. To achieve this, we propose an optimal selection framework which activates specific RF chains that amplify the digitally beamformed signals with the analogue beamforming network. Practical precoding is considered by including the effects of user interference noise, and hardware impairments in the system modelling.

THE ROYAL SOCIETY
PUBLISHING

1. Introduction

There is currently a strong requirement to identify the technical needs and possible solutions that will transform the wireless connectivity ecosystem for a better connected society in the future. Fifth generation (5G) wireless standards will start to address the consumer demands and performance enhancements for mobile communication in the next five years [1]. Looking forward, the Cisco data traffic report indicates that in 2023 video applications will generate up to 82% of the total mobile data traffic, with up to 29.3 billion networked devices [2]. The Ericsson mobility report [3] also forecasts that there will be 2.6 billion mobile connections by the end of 2025 and at least 45% of the world's population should be able to access 5G services in the same year. The next generation services are expected to be commercially implemented on a large scale in the next few years, e.g. in North America and North East Asia significant 5G subscriptions are expected to grow rapidly [3]. Future wireless systems require high data rates/throughput, improved coverage, lower latency, high mobility and reliability, and lower infrastructure costs [4,5].

One of the building blocks for fulfilling the requirements of the next generation mobile communications is the use of multiple-input multiple-output (MIMO) techniques in the form of large scale antenna arrays. Growth in spectrum availability will enhance overall network capacity in order to accommodate a large number of mobile users worldwide. Most current mobile broadband systems operate at frequencies below 6 GHz, but this spectrum is becoming increasingly crowded, especially in major cities and areas of high population density. The demand for additional spectrum can be fulfilled by the use of higher mmWave bands at carrier frequencies around 30-300 GHz [6,7]. There are many potential applications associated with mmWave communications, including fixed broadband access to the home, small cell communications in dense urban areas and vehicle-to-vehicle communications. However, moving up in carrier frequency leads to new challenges of higher path loss, more significant blocking effects and unconventional channel characteristics [8]. Consideration is now also being given to using the Terahertz spectrum at 300 GHz and above, where the channel conditions are often more severe, typically limiting the potential use cases to high data rate but very short range communications applications [9]. The use of MIMO technology can provide performance gains to help to mitigate the adverse channel effects, but these systems have hardware and power consumption constraints. The very name mmWave highlights the very small wavelengths associated with these frequencies which allows a large number of antennas to be placed on a compact space. Using a dedicated radio frequency (RF) chain for each antenna element would lead to the best data-rate performance. However, this solution is difficult to be implemented in practice because of the excessive power consumption and hardware complexity that results. Also, using wide bandwidth analogue to digital converters (ADCs) and digital to analogue converters (DACs) as part of the RF–chains at mmWave frequencies becomes a further source of hardware complexity and high power consumption. A parsimonious and energy efficient transceiver architecture is thus desired.

Figure 1 shows the multiuser *hybrid beamforming architecture* that is often studied in the literature for use at mmWave frequencies. The transmitter wishes to send K spatial streams to K receivers using spatial multiplexing techniques. Digital precoding is applied to the signals to direct these streams to the receiver using directional beamforming concepts. The digital precoder outputs are then converted into analogue form and amplified using L_{T} RF–chains. These waveforms are then directed to the transmitting antennas using an analogue precoder network. Various designs for the precoder network are possible, but may comprise a Butler matrix setup to generate fixed beam patterns or controllable phase shifters that allow dynamic beam patterns to be created. Typically the number of RF–chains L_{T} is much less than the number of transmitting antennas N_T because of their inherent broadband operation and high power consumption. The transmitting signal travels through the wireless channel which is different for each one of the K receiving terminals. We consider that each user equipment has N_{R} antennas and applies analogue RF combining to the received signal, before downconversion by a single RF–chain. The basic ideas of this architecture were discussed in [10], where the number of transmit and receive RF chains could be selected

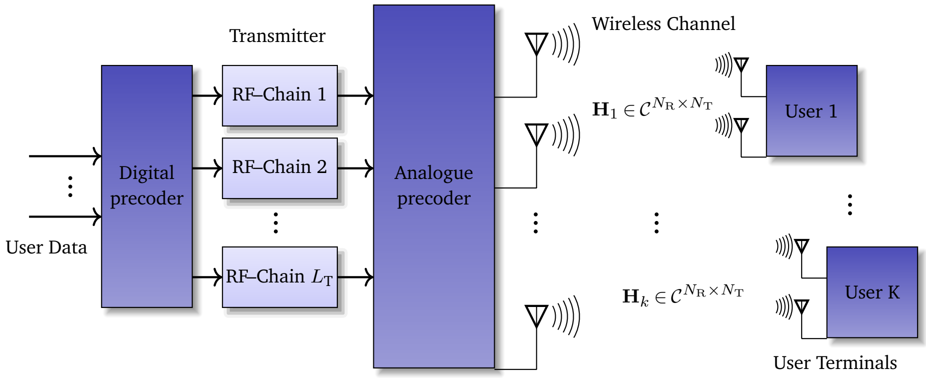

Figure 1. Block diagram of a hybrid precoding MIMO system with N_T transmitting antenna elements and L_T RF–chains. Each user terminal is equipped with N_R antennas.

according to the number of spatial streams that should be transmitted and received. Techniques for optimizing signal reception in this architecture are further explored in [11]. The concept of using sparse signal processing techniques to determine the best precoding and reception weights for maximising data throughput performance is described in [12]. A heuristic approach to hybrid beamforming that can achieve performance very close to the ideal case of using one RF–chain is also described in [13]. Detailed survey articles that discuss issues around mmWave communications and hybrid beamforming can be found in [14,15].

In recent years, optimizing the performance of the hybrid architecture has become the subject of intense study. The *fully-connected* architecture studied in [12] connects all of the antennas to each RF–chain, while the alternative *partially-connected* structure connects each RF–chain to only a subset of all antennas, which requires fewer phase shifters [16]. A detailed study in [17] also highlights that using the partially-connected setup can reduce RF losses in the system, improving performance. The partially-connected setup is therefore able to achieve a lower power consumption compared to the fully connected approach [18]. However, the partially-connected approach may suffer from increased co-channel interference, so a low-complexity interference cancellation precoding approach is proposed in [19]. Reference [20] studies the energy efficiency of a partially-connected HBF system, where each RF–chain is connected to only a subset of the available antennas. Other authors have tried to reduce the complexity through replacing some of the phase shifters with a network of RF switches, such as described in [21,22]. It is shown in these two papers that switches can operate at lower power consumption than phase shifters, enabling further energy savings with minimal impact on throughput performance. Energy efficient baseband signal processing methods to mitigate interference are also studied in [23]. A simple approach to receiver design is to make use of lens antennas in place of a phase shifters or switches and low complexity hardware implementations of this approach are reported in [24–26]. Reference [27] studies the impact of non-linear effects in transmitter power amplifiers and concludes that using only one RF–chain can be preferable in some scenarios to achieve the most energy efficient operating point of the system.

The power consumption can be reduced further though making use of low resolution quantisation of waveforms in MIMO transceivers. Jointly selecting low resolution quantisation at both the TX and the RX, and optimizing bit resolution with the precoding and combining designs can provide a highly energy efficient communication solution. This is due to the fact that the power consumption of DACs/ADCs scales exponentially with the number of bits used [31]. Reference [32] studies how the sampling bit resolution affects the achieved data rates using an additive quantisation noise model (AQNM) for the quantisation process. The AQNM approach is also used in [33] which shows the impact of low resolution sampling on the achieved data rate. The combination of a fully digital precoding TX with joint RF and baseband combining using low

Notations	Description
a	Scalar
$\mathbf{a}$	Vector
$\|\mathbf{a}\|_0$	l_0-norm of $\mathbf{a}$
$\mathbf{A}$	Matrix
$\|\mathbf{A}\|$	Determinant of $\mathbf{A}$
$\mathbf{A}^T$	Transpose of $\mathbf{A}$
$\mathbf{A}^H$	Complex conjugate transpose of $\mathbf{A}$
$\mathbf{A}^{(i)}$	i^{th} column of $\mathbf{A}$
$\|\mathbf{A}\|_F$	Frobenius norm of $\mathbf{A}$
$\mathcal{CN}(\mathbf{a};\mathbf{A})$	Complex Gaussian vector; mean $\mathbf{a}$, covariance $\mathbf{A}$
$\mathbb{R}^{A\times B}$	To represent a matrix of size $A \times B$ with real entries
$\mathbb{C}^{A\times B}$	To represent a matrix of size $A \times B$ with complex entries
$\mathcal{E}\{\cdot\}$	Expectation operator
$\mathbf{I}_N$	Identity matrix with size $N \times N$
$\mathbb{R}^+$	Set of positive real numbers
$\mathbb{R}\{\cdot\}$	Real part
$\text{tr}(\mathbf{A})$	Trace of $\mathbf{A}$
$\mathbf{X} \in \mathbb{C}^{A\times B}$	Complex-valued matrix $\mathbf{X}$ of size $A \times B$
$\mathbf{X} \in \mathbb{R}^{A\times B}$	Real-valued matrix $\mathbf{X}$ of size $A \times B$

Table 1. List of notations and their description.

resolution sampling at the RX is studied in [21]. Reference [34] proposes the idea of using a mixture of high and low resolution ADCs, which can achieve a higher EE than systems that use a fixed resolution ADCs at all receivers. The mmWave channel estimation problem when using low resolution sampling at the RX is also discussed in [35]. Care is needed when selecting the bit resolutions to be employed as the total power consumed may be dominated by a few ADCs or DACs operating at high resolution.

As noted above, the RF–chains can consume considerable power and increase the costs of the radio system [28], so another way to reduce energy is to optimize the number of activated RF–chains. A brute-force technique has been used in [29] to identify the most energy efficient hybrid precoder by designing the complete precoding solution for all of the choices for the number of RF–chains. A simpler alternative approach to optimizing the number of activated RF–chains was proposed in [30] which makes use of the Dinkelbach technique for optimizing the energy efficiency metric of data rate divided by power consumed.

This paper builds on the existing literature on optimal hybrid beamformer design [11] and the sparse solutions developed in [12]. More specifically, the paper builds on the Dinkelbach technique for energy efficient RF–chain selection developed in [30] for single user MIMO channels. We extend this work in two aspects. Firstly, we show how this method can be implemented in a multiuser broadcast channel where one transmitter uses beamforming to send multiple data streams simultaneously to multiple user terminals. Secondly, we show how our approach can take into account the impact of hardware imperfections within the precoding network. This requires a completely new approach based on the mathematical technique of *convex relaxation*, as compared to the simpler methods used in [30]. This is necessary to handle the increased complexity of selecting RF chains for the multi-user scenario. Further modifications are required to model hardware imperfections in the system and to account for the potential co-channel interference between multiple users.

Notations and Organisation

Table 1 provides a list of notations used in this paper along with their description.

The remainder of the paper is structured as follows: Section 2 reviews the literature on dynamic hybrid beamforming architectures. Section 3 describes the system and channel models. Section 4 discuss the EE maximisation problem where spectral efficiency and power consumption

Channel Attributes	Values
Bandwidth	100 MHz - 2 GHz
Base Station (BS) Antennas	64 - 256
Mobile Station (MS) Antennas	4 - 16
Channel Sparsity	High
Spatial Correlation	High
Angular Spread	<50 degrees
Orientation Sensitivity	High

Table 2. Channel attributes and their typical values for mmWave communication.

models are defined. In Section 5, we introduce the proposed RF selection algorithm. Section 6 presents simulation results to show the performance improvements and finally Section 7 presents conclusions to the paper.

2. System and Channel Model

(a) MmWave Channel

Making use of the very wide bandwidth channels available in the mmWave frequency bands is an important way to meet the needs of mobile broadband users in the next decade [36–38]. The higher path losses associated with mmWave spectrum compared to microwave bands can be mitigated through beamforming gains. These arise from using directional transmission and reception with large scale antenna arrays, i.e., MIMO systems. In addition, mmWave signal propagation is significantly affected by blockage effects, e.g., from the human body (attenuation from 20 to 35 dB [39]) and building materials such as brick (attenuation of 40 to 80 dB [40,41]). Table 2 discusses typical mmWave channel characteristics which are important attributes when considering mmWave frequency channels for next generation wireless standards. One very important property of a typical mmWave frequency channel is the high sparsity, i.e., there are only few significant propagation paths in the angle and delay domains [7,42].

Assuming that orthogonal frequency division multiplexing (OFDM) is being used, we make use of the flat fading channel model to model one subcarrier of a mmWave communication system. We consider P_k propagation paths for k^{th} user, where $k=1,\ldots,K$, N_{T} TX antennas and N_{R} RX antennas. The channel response matrix is given by:

$$\mathbf{H}_k = \sum_{p=1}^{P_k} \alpha_{k,p}\mathbf{a}(N_{\text{R}}, \theta_{k,p})\mathbf{b}^H(N_{\text{T}}, \phi_{k,p}) \tag{2.1}$$

The scalar $\alpha_{k,p}$ denotes the gain for the p^{th} multi-path component (MPC) for k^{th} user. The scalar P_k is the total number of MPCs for the k^{th} user and the vectors $\mathbf{b}(\cdot)\in\mathbb{C}^{N_{\text{T}}\times 1}$ and $\mathbf{a}(\cdot)\in\mathbb{C}^{N_{\text{R}}\times 1}$ are the steering vectors for the TX and RX respectively. Both TX and RX are assumed to use a uniform linear array (ULA) with antenna array spacing $d=\lambda/2$, so that the RX steering vector is defined as:

$$\mathbf{a}(N_{\text{R}}, \theta) = [1, e^{\text{j}\pi\cos(\theta)}, e^{\text{j}2\pi\cos(\theta)}, \ldots, e^{\text{j}(N-1)\pi\cos(\theta)}]^T. \tag{2.2}$$

The TX steering vector $\mathbf{b}(N_{\text{T}}, \phi)\in\mathbb{C}^{N_{\text{T}}\times 1}$ is defined similarly for the TX, with $\theta,\phi\in[-\frac{\pi}{2},\frac{\pi}{2}]$ representing the steering angle of arrival (AoA) and departure (AoD), respectively.

(b) System Model

We consider a mmWave downlink multi-user scenario, where the BS is equipped with an N_{T}-element ULA and serves K users. As shown in Figure 1, the BS employs analogue precoding represented by the matrix $\mathbf{F}_{\mathrm{RF}} \in \mathbb{C}^{N_\mathrm{T} \times L_\mathrm{T}}$. To capture the hardware imperfections of the analogue beamformer, we adopt the following linear model:

$$\mathbf{F}_{\mathrm{RF}} = \mathbf{F}_{\mathrm{RF}}^{\mathrm{ideal}} + \boldsymbol{\Phi}, \tag{2.3}$$

The matrix $\mathbf{F}_{\mathrm{RF}}^{\mathrm{ideal}} \in \mathcal{F}^{N_\mathrm{T} \times L_\mathrm{T}}$ where $\mathcal{F}^{N_\mathrm{T} \times L_\mathrm{T}}$ is the set of $N_\mathrm{T} \times L_\mathrm{T}$ matrices with constant modulus entries, given by $[\mathbf{F}_{\mathrm{RF}}]_{i,k} = e^{\mathrm{j}\pi(i-1)(k-1)/N}$. The error matrix $\boldsymbol{\Phi} \in \mathbb{C}^{N_\mathrm{T} \times L_\mathrm{T}}$ due to hardware imperfections is composed by identically independent distributed entries, $[\boldsymbol{\Phi}]_{i,k} \sim \mathcal{CN}(0, \sigma_\phi^2)$, and σ_ϕ^2 is the variance. The analogue precoder is connected to the baseband signals via L_T RF–chains. On the user side, each user equipment has an N_R-element ULA with a single RF–chain output.

The discrete-time received signal at k^{th} user terminal, with $k = 1, \ldots, K$, is expressed as:

$$y_k = \mathbf{w}_k^H \mathbf{H}_k \mathbf{F}_{\mathrm{RF}}^{\mathrm{ideal}} \mathbf{P}_{\mathrm{TX}}^{\frac{1}{2}} \mathbf{x} + \mathbf{w}_k^H \mathbf{H}_k \boldsymbol{\Phi} \mathbf{P}_{\mathrm{TX}}^{\frac{1}{2}} \mathbf{x} + \mathbf{w}_k^H \sum_{\ell=1, \ell \neq k}^{K-1} \mathbf{H}_\ell \mathbf{F}_{\mathrm{RF}} \mathbf{P}_{\mathrm{TX}}^{\frac{1}{2}} \mathbf{x} + n_k, \tag{2.4}$$

where:

- $\mathbf{x} \in \mathbb{C}^{L_\mathrm{T} \times 1}$ is the baseband signal for amplification and upconversion by the L_T RF–chains,
- $\mathbf{F}_{\mathrm{RF}} \in \mathbb{C}^{L_\mathrm{T} \times N_\mathrm{T}}$ is the analogue beamforming matrix, which considered to contain *noise* due to hardware imperfections and fluctuations in circuit behaviour.
- $\mathbf{P}_{\mathrm{TX}} \in \mathbb{R}^{L_\mathrm{T} \times L_\mathrm{T}}$ is a diagonal matrix whose entries contain the power amplification factors of the L_T RF–chains,
- $\mathbf{H}_k \in \mathbb{C}^{N_\mathrm{T} \times N_\mathrm{R}}$ is the mmWave channel response array between the k^{th} user and the BS,
- $\mathbf{w}_k \in \mathbb{C}^{N_\mathrm{R} \times 1}$ is the combiner beamforming vector applied for the k^{th} user: it is computed as the steering vector $\mathbf{b}$ that maximises the received signal power,
- $\mathbf{w}_k^H \mathbf{H}_k \boldsymbol{\Phi} \mathbf{x}$ is the noise due to hardware imperfections,
- $\mathbf{w}_k^H \sum_{\ell=1, \ell \neq k}^{K-1} \mathbf{H}_\ell \mathbf{F}_{\mathrm{RF}} \mathbf{x}$ is the inter-user interference, the vector $\mathbf{x} \in \mathbb{C}^{K \times 1}$ represents the broadcast signal, and
- $n_k \in \mathbb{C}$ represents the additive white Gaussian noise (AWGN) which is complex Gaussian distributed with zero-mean and variance σ_n^2, i.e., $n_k \sim \mathcal{CN}(0, \sigma_n^2)$.

3. Energy Efficiency Maximisation

The EE is defined as the ratio of the SE R (bit/sec/Hz) and the power P (Watts) [43],

$$\mathrm{EE} \triangleq \frac{R}{P} \quad \text{(bit/Hz/Joule)}. \tag{3.1}$$

Essentially, maximisation of the EE aims for simultaneous maximisation of the SE and minimisation of the required power. This problem can be expressed by the following contrained optimisation :

$$\max \frac{R}{P} \text{ subject to } R \geq R_{\min} \;\&\; P \leq P_{\max}, \tag{3.2}$$

where $R_{\min}$ and $P_{\max}$ are the predefined lower and upper bounds for the SE and power, respectively. The SE R and power P can be expressed as functions of several parameters, e.g., the analog/digital beamforming matrices, the number of RF–chains, for the TX and the RX respectively. Thus, equation (3.2) represents a fractional optimisation problem, where in general there may be no closed form expression for the solution [43]. Depending on the optimisation variable we choose to focus on, the SE R and the power P can be non-convex functions.

Let us consider the case where the optimisation variable is the power consumed by L_T RF–chains. Mathematically, the required energy for the operation of the i^{th} RF–chain can be represented by the i^{th} entry of the vector $\mathbf{p} \triangleq [p_1, \ldots, p_{L_\mathrm{T}}]^T$, where p_i is a positive real number, i.e., $p_i \in \mathbb{R}^+$. Furthermore, the overall consumed power P for the Hybrid BF MIMO system is composed by the terms:

$$P = P_{\text{amp}} + P_{\text{RFchains}} + P_{\text{circuit}} \text{ (Watts)}, \tag{3.3}$$

where P_{amp} is the power required by the amplifiers for signal transmission, P_{RFchains} is the consumed power at all the RF–chains and P_{circuit} is the consumed power at the digital and analogue circuit components. The consumed power for the RF–chains contributes significantly to the overall power P. In this work we focus on the minimisation of the RF–chains power, which is expressed as:

$$P_{\text{RFchains}} \triangleq \sum_{i=1}^{L_{\text{T}}} p_i. \text{ (Watts)} \tag{3.4}$$

Note that (3.4) is a convex function of the $\mathbf{p}$ since $\sum_{i=1}^{L_{\text{T}}} p_i = \|\mathbf{p}\|_1$.

For hybrid beamforming, we begin with a vector $\mathbf{s} \in \mathbb{C}^{K\times 1}$ which contains the K data streams for transmission: $\mathbf{s}$ obeys the property $E[\mathbf{s}\mathbf{s}^H] = \mathbf{I}_K$. The vector $\mathbf{s}$ is then pre-multiplied by the matrix $\mathbf{F_{BB}} \in \mathbb{C}^{L_{\text{T}}\times K}$ which represents the digital beamforming matrix at the transmitter to generate the baseband signal vector $\mathbf{x}$. The precoder is therefore decomposed as:

$$\mathbf{F} \triangleq \mathbf{F}_{\text{RF}}\mathbf{P}_{\text{TX}}^{\frac{1}{2}}\mathbf{F}_{\text{BB}}, \tag{3.5}$$

where $\mathbf{P}_{\text{TX}} \triangleq \text{diag}(\mathbf{p}_{\text{TX}})$ is the diagonal matrices representing the power consumption of the RF–chains at the TX, $\mathbf{p}_{\text{TX}}$. For the hybrid combiner that represents the signals processed at the receivers, we define the $N_R \times K$ matrix:

$$\mathbf{W} \triangleq [\mathbf{w}_1, \mathbf{w}_2, \ldots \mathbf{w}_K]. \tag{3.6}$$

The SE of the whole hybrid system is given by [12]:

$$R(\mathbf{P}_{\text{TX}}) = \log_2 |\mathbf{I}_K + \frac{1}{\sigma_{\text{n}}^2}\mathbf{Q}\mathbf{Q}^H|, \text{ (bit/sec/Hz)} \tag{3.7}$$

where the matrix $\mathbf{Q} \in \mathbb{C}^{K\times K}$ is defined as:

$$\mathbf{Q} \triangleq \underbrace{\mathbf{W}^H}_{\text{RX side}} \mathbf{H} \underbrace{\mathbf{F}_{\text{RF}}\mathbf{P}_{\text{TX}}^{\frac{1}{2}}\mathbf{F}_{\text{BB}}}_{\text{TX side}}. \tag{3.8}$$

To simplify the system analysis, usually the design of TX beamforming and RX beamforming are treated separately [12]. Thus, the SE considering the TX beamformer $\mathbf{F}$ given a pre-defined RX beamformer matrix W is given by:

$$R(\mathbf{P}_{\text{TX}}) = \log_2 |\mathbf{I}_K + \frac{1}{\sigma_{\text{n}}^2}\mathbf{W}^H\mathbf{H}\mathbf{F}_{\text{RF}}\mathbf{P}_{\text{TX}}^{\frac{1}{2}}\mathbf{F}_{\text{BB}}\mathbf{F}_{\text{BB}}^H\mathbf{P}_{\text{TX}}^{\frac{1}{2}}\mathbf{F}_{\text{RF}}^H\mathbf{H}^H\mathbf{W}|. \tag{3.9}$$

Maximising the EE ratio in equation (3.1) is in general a difficult mathematical problem. In order to make progress, we use the following theorem to study the mathematical properties of equation (3.9).

Theorem 1. *The SE given by (3.9) is a concave function of the diagonal matrix $\mathbf{P}_{\text{TX}}$, assuming $\mathbf{F}_{\text{BB}}\mathbf{F}_{\text{BB}}^H = \mathbf{I}_K$.*

Proof. Since $\mathbf{F}_{\text{BB}}\mathbf{F}_{\text{BB}}^H = \mathbf{I}_K$, the expression in (3.9) is written as:

$$\begin{aligned} R(\mathbf{P}_{\text{TX}}) &= \log_2 |\mathbf{I}_K + \frac{1}{\sigma_{\text{n}}^2}\mathbf{W}^H\mathbf{H}\mathbf{F}_{\text{RF}}\mathbf{P}_{\text{TX}}\mathbf{F}_{\text{RF}}^H\mathbf{H}^H\mathbf{W}| \\ &= \log_2 |\mathbf{I}_K + \frac{1}{\sigma_{\text{n}}^2}\mathbf{P}_{\text{TX}}\mathbf{F}_{\text{RF}}^H\mathbf{H}^H\mathbf{W}\mathbf{W}^H\mathbf{H}\mathbf{F}_{\text{RF}}| \\ &= \sum_{k=1}^{K} \log_2\left(1 + \frac{1}{\sigma_n^2}p_{\text{TX,q}}\lambda_q\right), \end{aligned} \tag{3.10}$$

where λ_k is the k^{th} eigenvalue of the matrix $(\mathbf{F}_{\text{RF}}^H \mathbf{H}^H \mathbf{W}\mathbf{W}^H \mathbf{H}\mathbf{F}_{\text{RF}})$. Recall that the sum of concave functions $\log_2(1+x)$ is also concave. Thus, $R(\mathbf{P}_{\text{TX}})$ is a concave function of $\mathbf{P}_{\text{TX}}$. □

Therefore, in our case, the fractional problem (3.2) is given as the ratio of a concave and a convex function. A common approach for solving fractional concave-convex problems is the Dinkelbach method (DB) [44], which replaces the fractional optimisation by an iterative sequence of simple problems based on the difference of the numerator and denominator. Specifically, the solution to the problem (3.2) is given successively by solving the problem:

$$\max(R_d - \kappa_d P_k) \text{ subject to } R_d \geq R_{\min} \text{ \& } P_d \leq P_{\max}, \tag{3.11}$$

for $d = 1, \dots D_{\max}$ where $D_{\max}$ is the maximum number of iterations of the method. In the above equation R_d and P_d are the SE and power for the d^{th} DM iteration,κ_d is the calculated EE ratio based on the previous estimation of R_{d-1} and P_{d-1}. Moreover, for concave-convex problems the DM method also provides convergence guarantees in order to find the globally best solution.

(a) EE maximisation via RF–Chain Subset Selection

Assigning a zero value to the power of the i^{th} RF–chain, p_i=0 represents the option of de-activating the corresponding RF–chain, so that it does not contribute to the overall power expenditure. However, due to the use of the zero value, the problem becomes a combinatorial one, where all possible combinations for the zero values that maximize the EE have to be exhaustively searched [30]. This means that the complexity of such an "exhaustive search" solution scales exponentially with the number of RF–chains L_{T}.

To overcome the issue of the non tractability of the exhaustive search, we consider the case where all RF–chains have equal power requirements, i.e., $p_i = p$. Then, the problem (3.9) can be formulated as a sparse subset selection one, by introducing a sparse RF–chain selection vector, s, with entries in the set $\{0, 1\}$. Incorporating this selection procedure into the expressions of rate in (3.9) and power in (3.4) for the TX, we have

$$R(\mathbf{S} = \log_2 |\mathbf{I}_K + \frac{1}{\sigma_{\text{n}}^2} \mathbf{W}^H \mathbf{H}\mathbf{F}_{\text{RF}} \mathbf{S}\mathbf{F}_{\text{BB}} \mathbf{F}_{\text{BB}}^H \mathbf{S}\mathbf{F}_{\text{RF}}^H \mathbf{H}^H \mathbf{W}|, \tag{3.12}$$

where $\mathbf{S} = \text{diag}(\mathbf{s})$, with $\mathbf{s} \triangleq [s_1, \cdots s_{L_{\text{T}}}] \in \{0,1\}^{L_{\text{TX}} \times 1}$, and $p_i = p$ for $i = 1, \dots, L_{\text{TX}}$. Note that (3.12), following the Theorem 1, is a concave function of $\mathbf{S}$. The power consumed by the RF–chains is expressed as:

$$P_{\text{RFchains}} \triangleq P_{\text{fix}} + p \sum_{i=1}^{L_{\text{T}}} s_i = P_{\text{fix}} + p \|\mathbf{S}\|_0, \text{ (Watts)} \tag{3.13}$$

where P_{fix} is a fixed power consumed by the system and which does not vary with the number of activated RF chains. The matrix $\mathbf{S}$ is a diagonal matrix where the value $s_i = 1$ denotes that the ith RF chain is activated, or set to zero otherwise. The introduction of the selection variable permits us the approximation of the combinatorial problem (3.11) with $s_i \in \{0,1\}$, $i = 1, \dots, L_{\text{T}}$, into an approximated convex problem with $s_i \in (0,1)$. Therefore, the d^{th} convex optimisation problem in (3.11) is transformed into a sparse subset selection, i.e.,

$$\min_{\mathbf{s}}(\kappa_d \|\mathbf{s}\|_1 - R_d(\mathbf{s})) \text{ subject to } R_d(\mathbf{s}) \geq R_{\min} \text{ \& } P_d(\mathbf{s}) \leq P_{\max}, \tag{3.14}$$

In [30] we provide an iterative algorithm that solves (3.14) via thresholding. Next, we will describe major modifications to the proposed approach that are required to optimize performance in the presence of hardware imperfections and multi-user co-channel interference. It will be seen that it can significantly reduce the overall complexity, as discussed in more detail in the following section.

4. Proposed Technique

The RF selection process can be seen as an additional block of the hybrid beamformer structure, that activates specific parts of the analogue beamformer. Thus, instead of finding the optimal phased-array matrix $\mathbf{F}_{\mathrm{RF}}$, we choose only a subset of columns of this matrix in order to achieve the highest EE. This selection module is added between the digital and analogue parts, and can be implemented by a network of switches. Essentially, it selects a subset of the L_{T} columns from a fixed codebook matrix which represents how the digital signals are forwarded to the analogue signal processing network.

Let us describe the selection mechanism that represents the active/inactive RF–chains at the BS, which is based on the approaches described in Section 3 above. For this, we use the *binary* matrix $\mathbf{S} \in \{0,1\}^{N_{\mathrm{T}} \times N_{\mathrm{T}}}$ defined in Section 3. Specifically, $\mathbf{S}$ is a diagonal matrix whose entries are either zero or one, A physical interpretation of $\mathbf{S}$ is possible by considering that this matrix represents a switching network. This network activates only a maximum of L_{T} outputs of an *extended* analogue combiner $\mathbf{F}^{e}_{\mathrm{RF}} \in \mathcal{F}^{N_{\mathrm{T}} \times N_{\mathrm{T}}}$. This extension ensures that the analogue front-end has the same number of inputs and outputs, thus, the selection is possible from the entire analogue, noisy, codebook given by:

$$\mathbf{F}^{e}_{\mathrm{RF}} = \mathbf{F}^{e,\mathrm{ideal}}_{\mathrm{RF}} + \boldsymbol{\Phi}^{e}, \tag{4.1}$$

where $\mathbf{F}^{e,\mathrm{ideal}}_{\mathrm{RF}} \in \mathcal{F}^{N_{\mathrm{T}} \times N_{\mathrm{T}}}$ is the *extended* ideal analogue beamformer and $\boldsymbol{\Phi}^{e} \in \mathbb{C}^{N_{\mathrm{T}} \times N_{\mathrm{T}}}$ is the *extended* noise matrix that captures hardware imperfections in the analogue circuitry.

These hardware imperfections may arise from one or more of the following sources [45]:

- **Phase Noise:** The oscillators used in the transmitter and receiver RF–chains are not perfect sine waves, but rather their frequencies slowly drift over time, causing variations in the channel model of the wireless links.
- **Mutual Coupling:** The antennas in a uniform or rectangular array are often spaced by half a wavelength as in equation (2.2). This means the antennas can be subject to re-radiation of the transmitted or received signals, an effect called mutual coupling. This distorts the steering vector from its ideal form [46].
- **RF Hardware Imperfections:** In this case non-linearities in the RF amplifiers can cause the signal to deviate from the linear model described in equation (2.4). In addition, some RF combiner circuits, such as the Rotman lens for performing beamforming [26] can cause distortion or spillover of the desired beam patterns.
- **Beam Squint:** In this case, the signal is sufficiently broadband that the steering vector defined in equation (2.2) is no longer accurate [47] for all frequencies within the bandwidth of the signal. Instead the steering vector becomes a function of the subcarrier index in OFDM data transmission.

Incorporating the selection matrix, the system model of the proposed framework is expressed as:

$$y_k = \mathbf{w}_k^H \mathbf{H}_k \mathbf{F}^{e,\mathrm{ideal}}_{\mathrm{RF}} \mathbf{S} \mathbf{P}_{\mathrm{TX}}^{\frac{1}{2}} \mathbf{x} + \zeta_k(\mathbf{S}) + \eta_k(\mathbf{S}) + n_k, \tag{4.2}$$

where $y_k \in \mathbb{C}$ is the received signal of the k^{th} user, the hardware noise component is given by,

$$\zeta_k(\mathbf{S}) = \mathbf{w}_k^H \mathbf{H}_k \boldsymbol{\Phi}^{e} \mathbf{S} \mathbf{P}_{\mathrm{TX}}^{\frac{1}{2}} \mathbf{x}, \tag{4.3}$$

while $\eta_k(\mathbf{S})$ is the interference that affects the k^{th} user, given by:

$$\eta_k(\mathbf{S}) = \mathbf{w}_k^H \sum_{\ell \neq k} \mathbf{H}_\ell \mathbf{F}^{e}_{\mathrm{RF}} \mathbf{S} \mathbf{P}_{\mathrm{TX}}^{\frac{1}{2}} \mathbf{x}. \tag{4.4}$$

For each user, the theoretical average SE is given by:

$$R_k(\mathbf{S}) = \log_2 \mathcal{E} \left\{ 1 + \frac{|\mathbf{w}_k^H \mathbf{H}_k \mathbf{F}^{e,\mathrm{ideal}}_{\mathrm{RF}} \mathbf{S} \mathbf{P}_{\mathrm{TX}}^{\frac{1}{2}} \mathbf{x}|^2}{|\zeta_k(\mathbf{S}) + \eta_k(\mathbf{S}) + n_k|^2} \right\}, \ \text{(bit/sec/Hz)} \tag{4.5}$$

Algorithm 1 Exhaustive search algorithm

Input: $\mathbf{F}_{\mathrm{RF}}^{e,\mathrm{ideal}}, \sigma_{\zeta_k}^2, \sigma_{\eta_k}^2, \sigma_{n_k}^2, \mathbf{w}_k, \mathbf{H}_k, , \mathcal{C}$
Output: $\mathbf{S}_{\mathrm{opt}}$
1: **for** $i = 1, \ldots, |\mathcal{C}|$ **do**
2: Compute the EE for $\mathbf{S}^{(i)} \in \mathcal{C}$ via (4.5) and (3.13)
3: **end for**
4: Find $\mathbf{S}_{\mathrm{opt}}$=arg $\max_{\mathbf{S}^{(i)}}$ EE s.t. $P(\mathbf{S}^{(i)}) \leq P_{\max}$ & $\bar{R}(\mathbf{S}^{(i)}) \geq R_{\min}$

Algorithm 2 Iterative minimisation algorithm

Input: $\mathbf{F}_{\mathrm{RF}}, \mathbf{F}_{\mathrm{BB}}, L_{\mathrm{T}}$
Output: $\mathbf{S}_{\mathrm{iter}}$
1: **for** $L_{\mathrm{T}} = 1, 2, \ldots, L_{\mathrm{T}}$ **do**
2: Compute the EE for $\mathbf{S}_{L_{\mathrm{T}}} = \begin{bmatrix} \mathbf{I}_{L_{\mathrm{T}}} & \mathbf{0}_{L_{\mathrm{T}}-L_{\mathrm{T}}} \\ \mathbf{0}_{L_{\mathrm{T}}-L_{\mathrm{T}}} & \mathbf{0}_{L_{\mathrm{T}}-L_{\mathrm{T}}} \end{bmatrix}$ via (4.5) and (3.13)
3: **end for**
4: Find $\mathbf{S}_{\mathrm{iter}} = \arg\max_{\mathbf{S}_{L_{\mathrm{T}}}}$ EE s.t. $P(\mathbf{S}_{L_{\mathrm{T}}}) \leq P_{\max}$

where the expectation $\mathcal{E}\{\cdot\}$ is performed over the joint space of $\{\mathbf{x}, \zeta_k, \eta_k\}$. Note that we assume that these noise sources are statistically independent, thus, there are no cross-correlation terms among them. Next, we provide an upper bound for $R_k(\mathbf{S})$ which is expressed based on the known covariance matrices of $\zeta_k(\mathbf{S})$ and $\eta_k(\mathbf{S})$.

In this work, we focus on the power that is consumed by each RF–chain, P_{RF}. Each RF–chain has a number of power consuming components, such as the ADC/DAC and power amplifiers. Thus, by activating only a subset of the RF–chains, the required power decreases significantly. This power level is computed using the model described in equation (3.13).

The EE problem for the multi-user downlink scenario, can be expressed as:

$$\max_{\mathbf{S}\in\mathcal{S}} \frac{\bar{R}(\mathbf{S})}{P(\mathbf{S})} \text{ s.t. } P(\mathbf{S}) \leq P_{\max} \text{ and } \bar{R}(\mathbf{S}) \geq R_{\min}, \tag{4.6}$$

where $\mathcal{S}$ is the set of the feasible diagonal matrices $\mathbf{S}$ which satisfy:

- $[\mathbf{S}]_{i,i} \in \{0, 1\}$, $[\mathbf{S}]_{i,k} = 0$ for $i \neq k$, and
- $\|\mathbf{S}\|_0 \leq L_{\mathrm{T}}$,

while $\bar{R}(\mathbf{S}) \triangleq \sum_{k=1}^{K} R_k(\mathbf{S})$. Due to the requirement $\mathbf{S} \in \{0, 1\}^{N_{\mathrm{T}} \times N_{\mathrm{T}}}$, problem (4.6) defines an integer concave-convex fractional problem, which is computationally very expensive to solve. The optimal subset can be obtained via exhaustive search over all possible combinations. Let the set $\mathcal{C}$ represents all possible combinations for the state (active/inactive) of the virtual switches for the L_{T} RF–chains. Then, the exhaustive search algorithm has to compute the EE of the i^{th} iteration, defined as

$$\mathrm{EE} \triangleq \frac{\bar{R}(\mathbf{S}^{(i)})}{P(\mathbf{S}^{(i)})}, \tag{4.7}$$

for all combinations $|\mathcal{C}|$ with $P(\mathbf{S}^{(i)}) \leq P_{\max}$ and $\bar{R}(\mathbf{S}^{(i)}) \geq R_{\min}$, and select the one with the highest EE. The exhaustive search method is summarized in Algorithm 1, where the remaining parameters are defined in Proposition 1 below. Note that, the number of the combinations $|\mathcal{C}|$ increases exponentially with L_{T}.

Before proceeding with the proposed technique, we would like to describe a sub-optimal, but computationally affordable and straightforward approach of solving (4.6). This could be implemented by selecting the minimum number of RF–chains, where the selection is performed by a naive technique, e.g., randomly or consecutively [30]. This approach can be implemented

Algorithm 3 Dinkelbach iterations
1: **for** $i = 1, 2, \ldots, I_{\max}$ **do**
2: $\kappa^{(i)} = \bar{R}^{(i)} / P^{(i)}$
3: Obtain $\mathbf{S}^{(i)}$ by solving (4.11)
4: Calculate $\bar{R}^{(i)}$ and $P^{(i)}$
5: **end for**

via a simple iterative search. Specifically, at each iteration the number of RF–chains L_{T} would increase by one until it reaches the maximum value, $L_{\mathrm{T}} = 1, 2, \ldots, L_{\mathrm{T}}$. At each iteration the EE will be computed and at the end of the iterations we select the L_{T} which provides the maximum EE. This iterative search is summarized in Algorithm 2.

In order to extend [30] to handle hardware imperfections and multi-user scenarios, the complexity of selecting the the correct RF chains in (4.6) grows exponentially with the number of RF chains L_{T}. In this paper we adopt the convex relaxation strategy, where the integer values are replaced by the set of real numbers $[\mathbf{S}]_{\ell,\ell} \in (0,1)$ [48]. This approach reflects the actual system hardware and permits study of scenarios where the impairments that affect different RF chains are not symmetric. Moreover, to deal with a fractional cost function, we employ Dinkelbach iterations [49]. This method is an iterative and parametric algorithm, where a sequence of simpler problems can be shown to converge to the global solution of the overall fractional problem. Let $\kappa^{(i)} \in \mathbb{R}$, for $i = 1, 2, \ldots, I_{\max}$, then one iteration of the DB method can be written as:

$$\max_{\mathbf{S} \in \mathcal{S}} \left\{ \sum_{k=1}^{K} R_k(\mathbf{S}) - \kappa^{(i)} P(\mathbf{S}) \right\} \tag{4.8}$$

The notation $\mathcal{S}$ denotes the set of diagonal matrices that satisfy both constraints $P(\mathbf{S}) \leq P_{\max}$ and $\bar{R}(\mathbf{S}) \geq R_{\min}$. The DB iteration steps are summarized in Algorithm 3. Parameter $\kappa^{(m)}$ is defined as the previous iteration EE computation [43], i.e.,

$$\kappa^{(m)} = \bar{R}^{(i-1)} / P^{(i-1)}, \tag{4.9}$$

with $\kappa^{(0)} = 1$.

To provide a computationally tractable solution for (3.14), we use a completely different approach to [30] and derive a novel lower bound approximation for the SE expression $\sum_k R_k$.

Proposition 1. *Given that the covariance matrices* $\mathcal{E}\{[\boldsymbol{\Phi}^e]_{l_{\mathrm{T}}}([\boldsymbol{\Phi}^e]_{l_{\mathrm{T}}})^H\}$ *and* $\mathcal{E}\{[\mathbf{F}^e_{\mathrm{RF}}]_{l_{\mathrm{T}}}([\mathbf{F}^e_{\mathrm{RF}}]_{l_{\mathrm{T}}})^H\}$ *are known for* $l_{\mathrm{T}} = 1, \ldots, L_{\mathrm{T}}$*, the achievable average SE for the* k^{th} *user, which is given by* (4.5) *can be lower bounded by:*

$$R_k \geq \frac{\|\mathbf{w}_k^H \mathbf{H}_k \mathbf{F}_{\mathrm{RF}}^{e,\mathrm{ideal}} \mathbf{S} \mathbf{P}_{\mathrm{TX}}^{\frac{1}{2}}\|^2}{\sigma_n^2 + \sigma_{\zeta_k}^2 + \sigma_{\eta_k}^2}, \tag{4.10}$$

where $\sigma_{\zeta_k}^2 \triangleq \boldsymbol{\xi}_k^H \mathbf{M} \boldsymbol{\xi}_k$, $\sigma_{\eta_k}^2 \triangleq \sum_{p \neq k} \boldsymbol{\xi}_p^H \mathbf{N} \boldsymbol{\xi}_p$ *and* σ_n^2 *is the variance of the AWGN.*

Proof. The proof is presented briefly in the Appendix. □

Thus, problem (4.6) becomes:

$$\max_{\mathbf{S} \in \mathcal{S}} \left\{ \sum_{k=1}^{K} \frac{\|\mathbf{w}_k^H \mathbf{H}_k \mathbf{F}_{\mathrm{RF}}^{e,\mathrm{ideal}} \mathbf{S} \mathbf{P}_{\mathrm{TX}}^{\frac{1}{2}}\|^2}{\sigma_n^2 + \sigma_{\zeta_k}^2 + \sigma_{\eta_k}^2} - \kappa^{(i)} P(\mathbf{S}) \right\}, \tag{4.11}$$

where the denominator is a function of the selection matrix $\mathbf{S}$, namely

$$\sigma_{\zeta_k}^2 = \boldsymbol{\xi}_k^H \mathcal{E} \left\{ \boldsymbol{\Phi}^e \mathbf{S} \mathbf{P}_{\mathrm{TX}}^{\frac{1}{2}} (\mathbf{P}_{\mathrm{TX}}^{\frac{1}{2}})^H \mathbf{S} (\boldsymbol{\Phi}^e)^H \right\} \boldsymbol{\xi}_k = \boldsymbol{\xi}_k^H \mathbf{M} \boldsymbol{\xi}_k, \tag{4.12}$$

Setting	Value
Carrier Frequency	28 GHz
System Bandwidth	100 MHz
Delay Spread	10 ns
Angle Spread	10°
Base Station (BS) Antennas N_T	32
Mobile Station (MS) Antennas N_R	8
Number of Users K	$5-15$
Fading Type	Rayleigh Fading
Number of Subpaths P_k	Randomly chosen for each user with $\mathcal{U}(0, 15)$

Table 3. Default simulation settings for the results presented in this section.

and

$$\sigma_{\eta_k}^2 = \sum_{p \neq k} \boldsymbol{\xi}_p^H \mathcal{E}\left\{\mathbf{F}_{\mathrm{RF}}^e \mathbf{S} \mathbf{P}_{\mathrm{TX}}^{\frac{1}{2}} (\mathbf{P}_{\mathrm{TX}}^{\frac{1}{2}})^H \mathbf{S} (\mathbf{F}_{\mathrm{RF}}^e)^H\right\} \boldsymbol{\xi}_p = \sum_{p \neq k} \boldsymbol{\xi}_p^H \mathbf{N} \boldsymbol{\xi}_p. \tag{4.13}$$

Formulas to compute the expectations in equations (4.12) and (4.13) are given in equations (A 6) and (A 8) in the Appendix. However, equation (4.11) is still non-convex over $\mathbf{S}$. To address this issue, we use the Titu's lemma on the summation term, i.e.,

$$\sum_{k=1}^{K} \frac{\|\mathbf{w}_k^H \mathbf{H}_k \mathbf{F}_{\mathrm{RF}}^{e,\mathrm{ideal}} \mathbf{S} \mathbf{P}_{\mathrm{TX}}^{\frac{1}{2}}\|^2}{\sigma_n^2 + \sigma_{\zeta_k}^2 + \sigma_{\eta_k}^2} \geq \frac{\sum_{k=1}^{K} \|\mathbf{w}_k^H \mathbf{H}_k \mathbf{F}_{\mathrm{RF}}^{e,\mathrm{ideal}} \mathbf{S} \mathbf{P}_{\mathrm{TX}}^{\frac{1}{2}}\|^2}{\sum_{k=1}^{K} \left(\sigma_n^2 + \sigma_{\zeta_k}^2 + \sigma_{\eta_k}^2\right)}. \tag{4.14}$$

Using the lower bound of (4.14) in (4.11), and employing the DB approach, we can replace the fractional cost function with:

$$\max_{\mathbf{S} \in \mathcal{S}} \sum_{k=1}^{K} \|\boldsymbol{\omega}_k \mathbf{S}\|^2 - \sum_{k=1}^{K} \left(\sigma_n^2 + \sigma_{\zeta_k}^2 + \sigma_{\eta_k}^2\right) - \kappa^{(i)} P(\mathbf{S}), \tag{4.15}$$

where $\boldsymbol{\omega}_k \triangleq \mathbf{w}_k^H \mathbf{H}_k \mathbf{F}_{\mathrm{RF}}^{e,\mathrm{ideal}}$. Since (4.15) is convex over $\mathbf{S}$, standard interior-point methods and publicly available software-packages can be used to solve (4.15). Note that, even if problems (4.6) and (4.15) result in different solution matrices $\mathbf{S}$, their EE performance is almost identical, as shown through the simulations in the next section.

5. Simulation Results

In this section, we use MATLAB$^{\mathrm{TM}}$ computer simulation results to evaluate the performance of the proposed method. All the results are averaged over 500 Monte-Carlo realisations. Let us first define the parameters and the system characteristics. We assume that the transmitter employs hybrid A/D transmit beamforming with N_{T} antennas, while the number of RF–chains is $L_{\mathrm{T}} \leq N_{\mathrm{T}}$. Each transmission broadcasts a zero-mean random Gaussian vector with $\mathbf{x} \in \mathbb{C}^{N_{\mathrm{R}} \times 1}$ and $\mathcal{E}\{\mathbf{x}\mathbf{x}^H\} = \mathbf{I}_{N_{\mathrm{R}}}$. We assume ULAs at both TX and RX sides and operating over a 28 GHz outdoor mmWave channel [50]. The K users are distributed uniformly random around the BS with maximum distance 5 meters. Also, the MPC P_k for the k^{th} user is selected uniformly random over the set $[1, 15]$. To focus on the TX performance, we assume digital combining is performed at the user equipment, i.e., $\mathbf{w}_k$ is defined as the k^{th} column of the left orthonormal matrix, obtained by the singular value decomposition of the channel matrix $\mathbf{H}_k$. Default channel parameter settings are shown in Table 3.

For the evaluation of the proposed technique in terms of EE and SE performance, we have considered the following cases for the TX:

(i) **Digital BF**: digital beamforming architecture ($N_{\mathrm{T}} = L_{\mathrm{T}}$), which represents the optimum from the achievable SE perspective,

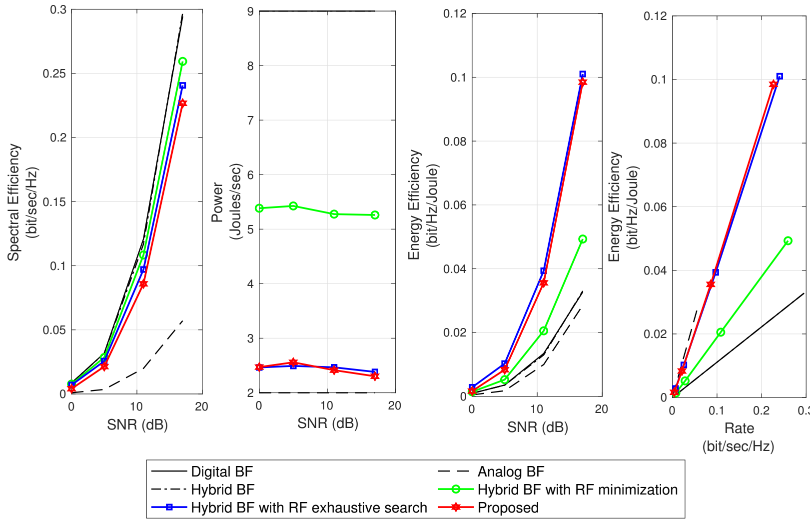

Figure 2. Performance comparisons over the transmit power P_{TX} for $N_{\mathrm{T}} = 8$, $L_{\mathrm{T}} = 8$, $N_{\mathrm{R}} = 8$. These results provide a sanity check for the convergence of the proposed technique with hardware impairments noise of, $\sigma_{\phi} = 0.01$. The number of the users was set to $K = 5$.

(ii) **Analogue BF**: analogue beamforming architecture with $L_{\mathrm{T}} = 1$, which represents a low power option with acceptable SE performance,

(iii) **Hybrid BF**: hybrid beamforming architecture with L_{T} RF–chains, where the beamforming matrices are obtained via [12]

(iv) **Iterative-HBF**: hybrid A/D transmit beamforming with the minimum number of RF–chains L_{T}, using Algorithm 2.

(v) **Exhaustive-HBF**: hybrid A/D transmit beamforming with the best subset of active RF–chains obtained by exhaustive search, using Algorithm 1. The results of this technique are limited by the number of RF–chains, i.e., $L_{\mathrm{T}} \leq 12$.

First, to perform a sanity check, we compare the results of the proposed technique and the Exhaustive-HBF for $N_{\mathrm{T}} = L_{\mathrm{T}} = N_{\mathrm{R}} = 8$. We keep the antenna arrays to small sizes due to the computational complexity of the Exhaustive-HBF. In Fig. 2 we plot the SE, Power, and EE with respect to the instantaneous signal-to-noise ratio (SNR), defined as:

$$\mathrm{SNR} = 10\log_{10}\left(\frac{1}{\sigma_n^2}\right), \tag{5.1}$$

as well as the EE versus SE. It can be verified that the proposed technique maximizes the EE, following closely the performance of the optimum Exhaustive-HBF algorithm. The Iterative-HBF cannot reach the EE of the proposed technique, since it minimizes the number of the RF–chains (e.g., the utilized codebook beams) and it does not consider the best subset of RF–chains to use. Note that the Hybrid BF and Digital BF have the same EE performance, since $N_{\mathrm{T}} = L_{\mathrm{T}}$. The Analogue BF method has only one RF–chain and thus it has a minimal number of digital components; although it has low power consumption, it exhibits the lowest EE. This comes from the fact that the transmitted signal needs to be multiplexed in time or frequency between the different user terminals, as the transmitter cannot achieve spatial multiplexing. The power of the proposed and exhaustive-HBF techniques is around 2.5 Joules/sec for all SNRs. This is half of the power required by the Iterative-HBF and the 1/4-th of the Digital and Hybrid BFs.

Next, we increase the antenna array size of the BS to $N_{\mathrm{T}} = 32$, while the number of the active RF–chains, that connect the analogue and the digital parts at the BS, remain $L_{\mathrm{T}} = 8$. Recall that, in order to focus on the performance at the BS, each user employs a digital combiner with $N_{\mathrm{R}} = 8$.

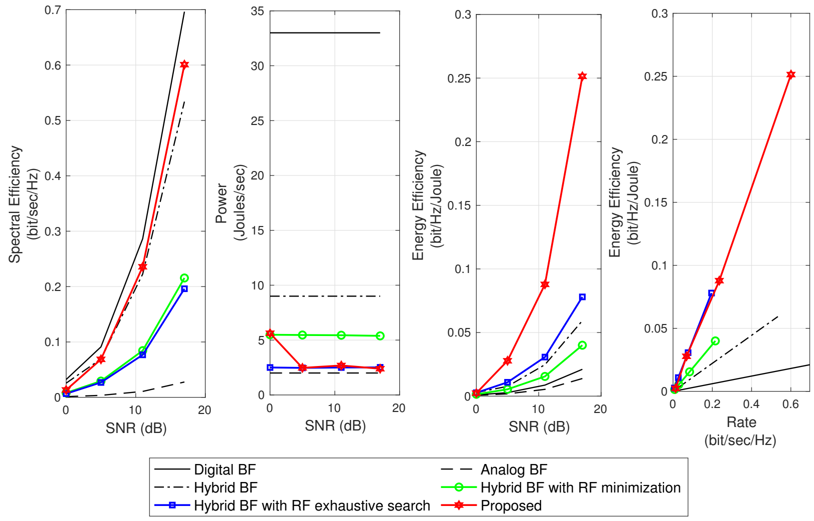

Figure 3. Performance comparisons over the transmit power P_{TX} for $N_{\mathrm{T}} = 32$, $L_{\mathrm{T}} = 8$, $N_{\mathrm{R}} = 8$ and $\sigma_{\phi} = 0.01$. The number of the users was set to $K = 5$.

In Fig. 3, we show the SE and EE with respect to the SNR for $N_{\mathrm{TX}} = 32$ and $\sigma_{\phi} = 0.01$. The proposed technique outperforms the other baselines, even the Exhaustive-HBF. This is possible, since it is able to search over the whole N_{T} codebook space, having polynomial computational complexity. To achieve this performance, the analogue part (e.g., phase-shifters) of the BS beamformer have been increased, based on the proposed design. Thus, a trade-off between the hardware complexity and the EE performance is possible. Note that in this work, we consider that the power consumption of each phase-shifter is negligible compared to the power consumption of each RF–chain [25]. This structure can be realized by using very energy efficient elements, e.g., passive phase-shifters or a Butler matrix [51]. The required power for the proposed and the Exhaustive-HBF techniques remain below 5 Joules/sec. The Iterative-HBF requires the double energy per second, while the Hybrid BF four times more. The results for this case indicate the high power consumption for the Digital BF, which is over 30 Joules/sec.

In Fig. 4 we plot the achievable SE, Power and EE over the number of RF–chains L_{T}. The proposed design is able to achieve superior EE performance when compared with the other hybrid BF techniques. The achievable SE of Iterative-HBF and Exhaustive-HBF is very similar. Recall that the Iterative-HBF selects the minimum number of RF–chains that achieves the best EE, while Exhaustive-HBF searches for the best overall subset of RF–chains. However, the search space of the latter is constrained to 8 codebook beams, due to the very high computational burden.

In Fig. 5 we show the achievable SE, Power and EE with respect to the number of the users K. The proposed technique achieves high SE, following the Hybrid BF and Digital BF curves. The Power consumption of the proposed technique remains at the same level with the Analog BF and Exhaustive-HBF, thus, it achieves the highest EE compared to the other approaches. This indicates that the proper design of the beamformer via RF–chain selection focus the beams to different locations.

It is important to note that, the proposed design has very similar SE performance with the Hybrid BF approach [12], as shown in Figs. 3 and 4. Indeed, a connection between the proposed technique and the greedy algorithm introduced in [12] exists. Specifically, the algorithm of [12] estimates a sparse vector which corresponds to the digital part of the beamformer. In the proposed design, we seek a sparse binary vector which is also part of the digital beamformer. Additionally, the analogue parts are designed using static analogue codebooks in both approaches. However, the analogue part of [12] is assumed to be drawn from an idealized discrete codebook, while, in the proposed technique we explicitly model the introduced noise due to hardware imperfections.

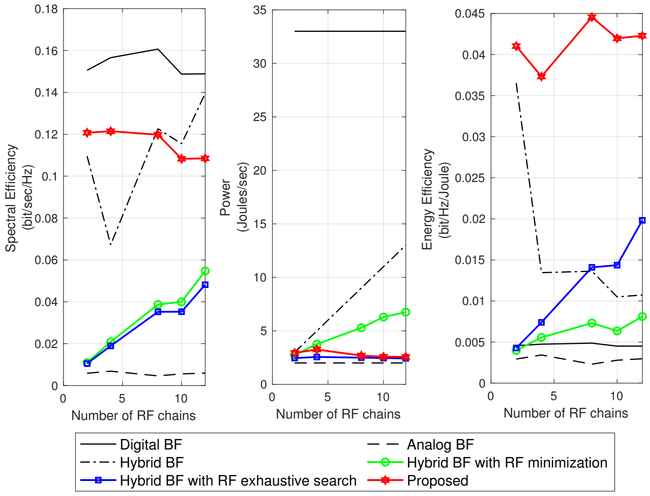

Figure 4. Spectral efficiency, Power and Energy efficiency over the number of RF chains L_{T}, for $N_{\mathrm{T}} = 32$, $N_{\mathrm{R}} = 8$, and $\sigma_{\phi} = 0.01$. The number of the users was set to $K = 8$.

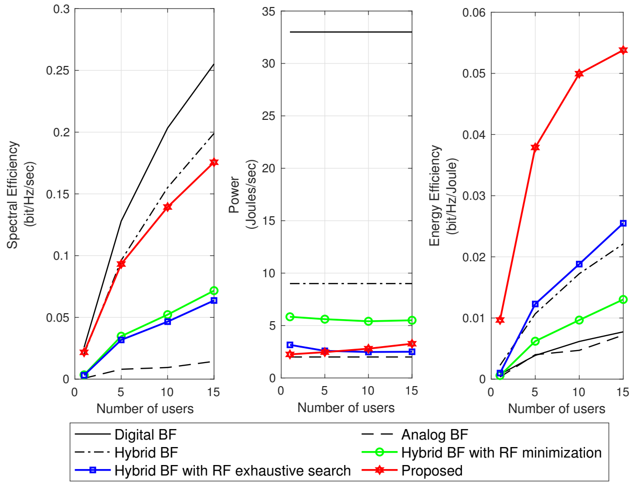

Figure 5. Spectral efficiency, Power and Energy efficiency over the number of users K, for $N_{\mathrm{T}} = 32$, $L_{\mathrm{T}} = 8$, $N_{\mathrm{R}} = 8$, and $\sigma_{\phi} = 0.01$.

Moreover, the proposed technique outperforms [12] in terms of EE, since via the switches, it deactivates parts of the analogue beamformer which do not contribute significantly to the overall SE performance.

6. Conclusion

This paper discusses the advantages of hybrid beamforming architectures for millimetre-wave wireless communications systems. By using a small number of radio frequency chains compared to the number of antennas, it is possible to improve the energy efficiency of communication. A novel radio frequency chain selection architecture is described to allocate the best predefined analogue codebook that maximizes the energy efficiency performance of the transmitter. Via simulation results, we showed that its beneficial in terms of EE, to activate a subset of radio frequency chains, rather than always to to use the maximum number that are available. The proposed algorithm outperforms all the baselines in terms of both energy and spectrum efficiency, when the transmitter has a large number of antenna terminals.

7. Acknowledgements

The authors gratefully acknowledge helpful discussions on the topic of this research with Dr. Christos Tsinos and Prof. Symeon Chatzinotas from the University of Luxembourg.

8. Funding

This work was supported by the UK EPSRC [grant numbers EP/P000703/1, EP/S026622/1].

Appendix

Starting from (4.5), the k^{th} user rate can be is lower bounded by

$$R_k(\mathbf{S}) = \log_2 \mathcal{E}\{1+\gamma_k(\mathbf{S})\} \geq \frac{\mathcal{E}\{\gamma_k(\mathbf{S})\}}{\mathcal{E}\{1+\gamma_k(\mathbf{S})\}} > c\mathcal{E}\{\gamma_k(\mathbf{S})\}, \tag{A 1}$$

where $\gamma_k(\mathbf{S}) \triangleq \frac{|\boldsymbol{\omega}_k^H \mathbf{S}\mathbf{P}_{\text{TX}}^{\frac{1}{2}}\mathbf{x}|^2}{|\zeta_k(\mathbf{S})+\eta_k(\mathbf{S})+n_k|^2}$ and $\boldsymbol{\omega}_k^H \triangleq \mathbf{w}_k^H \mathbf{H}_k \mathbf{F}_{\text{RF}}^{\text{e,ideal}}$. Using Titu's lemma we have that:

$$\mathcal{E}\left\{\frac{|\boldsymbol{\omega}_k^H \mathbf{S}\mathbf{P}_{\text{TX}}^{\frac{1}{2}}\mathbf{x}|^2}{|\zeta_k(\mathbf{S})+\eta_k(\mathbf{S})+n_k|^2}\right\} \geq \frac{\mathcal{E}\left\{|\boldsymbol{\omega}_k^H \mathbf{S}\mathbf{P}_{\text{TX}}^{\frac{1}{2}}\mathbf{x}|^2\right\}}{\mathcal{E}\{|\zeta_k(\mathbf{S})+\eta_k(\mathbf{S})+n_k|^2\}}. \tag{A 2}$$

Given that $\mathcal{E}\{\mathbf{x}\mathbf{x}^H\} = \mathbf{I}$ and $\mathbf{S}^2 = \mathbf{S}$, the numerator of (A 2) is given by:

$$\mathcal{E}\left\{|\boldsymbol{\omega}_k^H \mathbf{S}\mathbf{P}_{\text{TX}}^{\frac{1}{2}}\mathbf{x}|^2\right\} = \boldsymbol{\omega}_k^H \mathbf{S}\mathbf{P}_{\text{TX}}^{\frac{1}{2}} \mathcal{E}\{\mathbf{x}\mathbf{x}^H\}(\mathbf{P}_{\text{TX}}^{\frac{1}{2}})^H \mathbf{S}\boldsymbol{\omega}_k = \boldsymbol{\omega}_k^H \mathbf{P}_{\text{TX}}^{\frac{1}{2}} \mathbf{S}(\mathbf{P}_{\text{TX}}^{\frac{1}{2}})^H \boldsymbol{\omega}_k = \left\|\boldsymbol{\omega}_k^H \mathbf{S}\mathbf{P}_{\text{TX}}^{\frac{1}{2}}\right\|^2. \tag{A 3}$$

Given that the hardware noise, the user interference noise and the white Gaussian noise are mutually independent, the denominator of (A 2) is given by:

$$\mathcal{E}\{|\zeta_k(\mathbf{S})+\eta_k(\mathbf{S})+n_k|^2\} = \mathcal{E}\{|\zeta_k(\mathbf{S})|^2\} + \mathcal{E}\{|\eta_k(\mathbf{S})|^2\} + \mathcal{E}\{|n_k|^2\}, \tag{A 4}$$

where $\mathcal{E}\{|n_k|^2\} = \sigma_n^2$. This expectation can be easily calculated as $\mathbf{S}$ and $\mathbf{P}_{\text{TX}}$ are both diagonal matrices,

$$\begin{aligned} \mathcal{E}\{|\zeta_k(\mathbf{S})|^2\} &= \boldsymbol{\xi}_k^H \mathcal{E}\left\{\boldsymbol{\Phi}^e \mathbf{S}\mathbf{P}_{\text{TX}}^{\frac{1}{2}} \mathbf{x}\mathbf{x}^H (\mathbf{P}_{\text{TX}}^{\frac{1}{2}})^H \mathbf{S}(\boldsymbol{\Phi}^e)^H\right\} \boldsymbol{\xi}_k \\ &= \boldsymbol{\xi}_k^H \mathcal{E}\left\{\boldsymbol{\Phi}^e \mathbf{S}\mathbf{P}_{\text{TX}}^{\frac{1}{2}} (\mathbf{P}_{\text{TX}}^{\frac{1}{2}})^H \mathbf{S}(\boldsymbol{\Phi}^e)^H\right\} \boldsymbol{\xi}_k = \boldsymbol{\xi}_k^H \mathbf{M}\boldsymbol{\xi}_k, \end{aligned} \tag{A 5}$$

with $\boldsymbol{\xi}_k \triangleq \mathbf{w}_k^H \mathbf{H}_k$. The expectation present in the matrix $\mathbf{M}$ can be easily calculated as $\mathbf{S}$ and $\mathbf{P}_{\text{TX}}$ are both diagonal matrices. The lth row and mth column entry of this matrix can be calculated as:

$$\mathbf{M}_{l,m} = \mathcal{E}\left\{\boldsymbol{\Phi}_l^e \mathbf{S}\mathbf{P}_{\text{TX}}^{\frac{1}{2}} \mathbf{P}_{\text{TX}}^{\frac{1}{2}} \mathbf{S}(\boldsymbol{\Phi}_m^e)^H\right\} = \sum_{n=1}^{N_T} \mathbf{S}(n)\mathbf{P}_{\text{TX}}(n)\mathcal{E}\left\{\boldsymbol{\Phi}_{l,n}^e(\boldsymbol{\Phi}_{n,m}^e)^H\right\}. \tag{A 6}$$

Here $\boldsymbol{\Phi}_l^e$ denotes the lth row vector and $\boldsymbol{\Phi}_{l,n}^e$ denotes the lth row and nth column of $\boldsymbol{\Phi}^e$ respectively. The notation $\mathbf{S}(l)$ is the lth diagonal entry of $\mathbf{S}$ and $\mathbf{P}_{\text{TX}}(l)$ is the lth diagonal entry of $\mathbf{P}_{\text{TX}}$.

Similarly,

$$\begin{aligned}\mathcal{E}\left\{|\eta_k(\mathbf{S})|^2\right\} &= \sum_{p\neq k}|\boldsymbol{\xi}_p^H\mathbf{F}_{\mathrm{RF}}^e\mathbf{S}\mathbf{P}_{\mathrm{TX}}^{\frac{1}{2}}\mathbf{x}|^2 \\ &= \sum_{p\neq k}\boldsymbol{\xi}_p^H\mathcal{E}\left\{\mathbf{F}_{\mathrm{RF}}^e\mathbf{S}\mathbf{P}_{\mathrm{TX}}^{\frac{1}{2}}\mathbf{x}\mathbf{x}^H(\mathbf{P}_{\mathrm{TX}}^{\frac{1}{2}})^H\mathbf{S}(\mathbf{F}_{\mathrm{RF}}^e)^H\right\}\boldsymbol{\xi}_p \\ &= \sum_{p\neq k}\boldsymbol{\xi}_p^H\mathcal{E}\left\{\mathbf{F}_{\mathrm{RF}}^e\mathbf{S}\mathbf{P}_{\mathrm{TX}}^{\frac{1}{2}}(\mathbf{P}_{\mathrm{TX}}^{\frac{1}{2}})^H\mathbf{S}(\mathbf{F}_{\mathrm{RF}}^e)^H\right\}\boldsymbol{\xi}_p = \sum_{p\neq k}\boldsymbol{\xi}_p^H\mathbf{N}\boldsymbol{\xi}_p, \qquad (A\,7)\end{aligned}$$

where $\boldsymbol{\xi}_p^h \triangleq \mathbf{w}_p^H\mathbf{H}_p$, $\mathcal{E}\{\boldsymbol{\Phi}^e\mathbf{S}(\boldsymbol{\Phi}^e)^H\}$ and $\mathcal{E}\{\mathbf{F}_{\mathrm{RF}}^e\mathbf{S}(\mathbf{F}_{\mathrm{RF}}^e)^H\}$ are the covariance matrices of $\zeta_k(\mathbf{S})$ and $\eta_k(\mathbf{S})$ which depend on the selection matrix $\mathbf{S}$. By analogy to equation (A 6), we have for $\mathbf{N}$

$$\begin{aligned}\mathbf{N}_{l,m} &= \mathcal{E}\left\{\mathbf{F}_{\mathrm{RF}}^e(l)\mathbf{S}\mathbf{P}_{\mathrm{TX}}^{\frac{1}{2}}\mathbf{P}_{\mathrm{TX}}^{\frac{1}{2}}\mathbf{S}(\mathbf{F}_{\mathrm{RF}}^e(m))^H\right\} \\ &= \sum_{n=1}^{L_T}\mathbf{S}(n)\mathbf{P}_{\mathrm{TX}}(n)\mathcal{E}\left\{\mathbf{F}_{\mathrm{RF}}^e(l,n)(\mathbf{F}_{\mathrm{RF}}^e(n,m))^H\right\}, \qquad (A\,8)\end{aligned}$$

where $\mathbf{F}_{\mathrm{RF}}^e(l)$ denotes the lth row vector and $\mathbf{F}_{\mathrm{RF}}^e(l,n)$ is the lth row and nth column entry of $\mathbf{F}_{\mathrm{RF}}^e$, respectively. Using equations (A 3), (A 5) and (A 7) leads to the desired result.

References

1. C.-L. I., S. Han, Z. Xu, Q. Sun and Z. Pan, "5G: rethink mobile communications for 2020+," Phil. Trans. R. Soc. A., 2016.
2. Cisco visual networking index forecast 2017–22.
3. Ericsson Mobility Report, pp. 1-32, Nov. 2018.
4. NGMN 5G White Paper, pp. 1-125, Feb. 2015.
5. 2020 Networld White Paper for Research Beyond 5G, pp. 1-43, Oct. 2015.
6. Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101-107, June 2011.
7. S. Rangan, T. S. Rappaport and E. Erkip, "Millimeter-wave cellular wireless networks: potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.
8. IEEE 5G and Beyond Technology Roadmap Whitepaper, pp. 1-39, Oct. 2017.
9. B. Peng, K. Guan, A. Kuter, S. Rey, M. Patzold and T. Kuerner, "Channel Modeling and System Concepts for Future Terahertz Communications: Getting Ready for Advances Beyond 5G," in IEEE Vehicular Technology Magazine, vol. 15, no. 2, pp. 136-143, June 2020.
10. X. Zhang, A. F. Molisch and S.-Y. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4091-4103, Nov. 2005.
11. J. Ahmadi-Shokouh, S. H. Jamali and S. Safavi-Naeini, "Optimal receive soft antenna selection for MIMO interference channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 5893-5903, Dec. 2009.
12. O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499-1513, Mar. 2014.
13. F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE Journ. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501-513, Apr. 2016.
14. R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE Journ. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436-453, Apr. 2016.
15. I. Ahmed et al., "A Survey on Hybrid Beamforming Techniques in 5G: Architecture and System Model Perspectives," in IEEE Communications Surveys and Tutorials, vol. 20, no. 4, pp. 3060-3097, Fourth quarter 2018.
16. X. Yu, J. Shen, J. Zhang and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE Journ. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 485-500, Feb. 2016.

17. C. Lin and G. Y. Li, "Energy-Efficient Design of Indoor mmWave and Sub-THz Systems With Antenna Arrays," in IEEE Transactions on Wireless Communications, vol. 15, no. 7, pp. 4660-4672, July 2016.
18. C. G. Tsinos, S. Maleki, S. Chatzinotas and B. Ottersten, "On the energy-efficiency of hybrid analog-digital transceivers for single- and multi-carrier large antenna array systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1980-1995, Sept. 2017.
19. X. Gao, L. Dai, S. Han, C.-L. I and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Ar. Com.*, vol. 34, no. 4, pp. 998-1009, Apr. 2016.
20. E. Vlachos, A. Kaushik and J.S. Thompson, "Energy efficient transmitter with low resolution DACs for massive MIMO with partially connected hybrid architecture," *Veh. Tech. Conf. (VTC Spring), Porto, Portugal*, pp. 1-5, Jun. 2018.
21. R. Mendez-Rial, C. Rusu, N. Gonzalez-Prelcic, A. Alkhateeb and R. W. Heath, "Hybrid MIMO Architectures for Millimeter Wave Communications: Phase Shifters or Switches?," in IEEE Access, vol. 4, pp. 247-267, 2016.
22. S. Payami, M. Ghoraishi, M. Dianati and M. Sellathurai, "Hybrid Beamforming With a Reduced Number of Phase Shifters for Massive MIMO Systems," in IEEE Transactions on Vehicular Technology, vol. 67, no. 6, pp. 4843-4851, June 2018.
23. Q. Yu, C. Han, L. Bai, J. Choi and X. Shen, "Low-Complexity Multiuser Detection in Millimeter-Wave Systems Based on Opportunistic Hybrid Beamforming," in IEEE Transactions on Vehicular Technology, vol. 67, no. 10, pp. 10129-10133, Oct. 2018.
24. J. Brady, N. Behdad and A. M. Sayeed, "Beamspace MIMO for Millimeter-Wave Communications: System Architecture, Modeling, Analysis, and Measurements," in IEEE Transactions on Antennas and Propagation, vol. 61, no. 7, pp. 3814-3827, July 2013.
25. P. V. Amadori and C. Masouros, "Low RF-complexity millimeter-wave beamspace-MIMO systems by beam selection," IEEE Trans. Commun., vol. 63, no. 6, pp. 2212–2223, 2015.
26. M. A. Babar Abbasi, H. Tataria, V. F. Fusco and M. Matthaiou, "On the Impact of Spillover Losses in 28 GHz Rotman Lens Arrays for 5G Applications," 2018 IEEE MTT-S International Microwave Workshop Series on 5G Hardware and System Technologies (IMWS-5G), Dublin, 2018, pp. 1-3.
27. N. N. Moghadam, G. Fodor, M. Bengtsson and D. J. Love, "On the Energy Efficiency of MIMO Hybrid Beamforming for Millimeter-Wave Systems With Nonlinear Power Amplifiers," in IEEE Transactions on Wireless Communications, vol. 17, no. 11, pp. 7208-7221, Nov. 2018.
28. E. Bjornson, L. Sanguinetti, J. Hoydis and M. Debbah, "Optimal design of energy-efficient multi-user MIMO systems: is massive MIMO the answer?," *IEEE Trans. Wireless Commun.*, vol. 14, no.6, pp. 3059-3075, Jun. 2015.
29. R. Zi, X. Ge, J. Thompson, C. Wang, H. Wang and T. Han, "Energy efficiency optimization of 5G radio frequency chain systems," *IEEE Journ. Sel. Areas Commun.*, vol. 34, no. 4, pp. 758-771, Apr. 2016.
30. A. Kaushik, J. Thompson, E. Vlachos, C. Tsinos and S. Chatzinotas, "Dynamic RF Chain Selection for Energy Efficient and Low Complexity Hybrid Beamforming in Millimeter Wave MIMO Systems," *IEEE Trans. Green Commun. Netw.*, vol. 3, no. 4, pp. 886-900, Dec. 2019.
31. B. Murmann, "The Race for the Extra Decibel: A Brief Review of Current ADC Performance Trajectories," *IEEE Solid-State Circuits Mag.*, vol. 7, no. 3, pp. 58-66, 2015.
32. O. Orhan, E. Erkip and S. Rangan, "Low power analog-to-digital conversion in millimeter wave systems: Impact of resolution and bandwidth on performance," *2015 Info. Theory Appl. Workshop, San Diego, CA*, pp. 191-198, 2015.
33. L. Fan, S. Jin, C. Wen and H. Zhang, "Uplink achievable rate for massive MIMO systems with low-resolution ADC," *IEEE Commun. Letters*, vol. 19, no. 12, pp. 2186-2189, Oct. 2015.
34. J. Zhang, L. Dai, Z. He, S. Jin and X. Li, "Performance analysis of mixed-ADC massive MIMO systems over Rician fading channels," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1327-1338, June 2017.
35. A. Kaushik, E. Vlachos, J. Thompson and A. Perelli, "Efficient channel estimation in millimeter wave hybrid MIMO systems with low resolution ADCs", *IEEE EUSIPCO, Rome, Italy*, Sept. 2018.
36. J. G. Andrews et al., "What will 5G be?," *IEEE Journ. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065-1082, June 2014.
37. T. S. Rappaport et al., "Millimeter wave mobile communications for 5G cellular: it will work!," *IEEE Access*, vol. 1, pp. 335-349, 2013.

38. F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.
39. J. S. Lu, D. Steinbach, P. Cabrol, and P. Pietraski , "Modeling human blockers in millimeter wave radio links," *ZTE Commun.*, vol. 10, no. 4, pp. 23-28, Dec. 2012.
40. A. V. Alejos, M. G. Sanchez and I. Cuinas, "Measurement and analysis of propagation mechanisms at 40 GHz: Viability of site shielding forced by obstacles," *IEEE T. Veh. Tech.*, vol. 57, no. 6, pp. 3369-3380, Nov. 2008.
41. H. Zhao et al., "28 GHz millimeter wave cellular communication measurements for reflection and penetration loss in and around buildings in New York City," *IEEE Int. Conf. Commun. (ICC)*, pp. 5163-5167, 2013.
42. M. Akdeniz, et al., "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, pp. 1164-1179, June 2014.
43. A. Zappone and E. Jorswieck, "Energy Efficiency in Wireless Networks via Fractional Programming Theory," in *Energy Efficiency in Wireless Networks via Fractional Programming Theory*, Now Foun. Trends, 2015.
44. W. Dinkelbach, "On nonlinear fractional programming," *Management Science*, vol. 13, no. 7, pp. 492-498, Mar. 1967.
45. X. Yang, M. Matthaiou, J. Yang, C. Wen, F. Gao and S. Jin, "Hardware-Constrained Millimeter-Wave Systems for 5G: Challenges, Opportunities, and Solutions," in IEEE Communications Magazine, vol. 57, no. 1, pp. 44-50, January 2019.
46. X. Artiga, B. Devillers and J. Perruisseau-Carrier, "Mutual coupling effects in multi-user massive MIMO base stations," Proceedings of the 2012 IEEE International Symposium on Antennas and Propagation, Chicago, IL, 2012, pp. 1-2.
47. M. Wang, F. Gao, N. Shlezinger, M. F. Flanagan and Y. C. Eldar, "A Block Sparsity Based Estimator for mmWave Massive MIMO Channels With Beam Squint," in IEEE Transactions on Signal Processing, vol. 68, pp. 49-64, 2020.
48. D. Palomar and Y. Eldar, "Convex Optimization in Signal Processing and Communications," Cambridge: Cambridge University Press, 2009.
49. J. P. Crouzeix and J. A. Ferland, "Algorithms for generalized fractional programming," Math. Program., vol. 52, no. 1–3, pp. 191–207, May 1991.
50. A. Forenza, D. J. Love, and R. W. Heath, "Simplified spatial correlation models for clustered MIMO channels with different array configurations," IEEE Trans. Veh. Technol., vol. 56, no. 4 II, pp. 1924–1934, Jul. 2007.
51. J. Li, L. Xiao, X. Xu, X. Su, and S. Zhou, "Energy-efficient Butler-matrix-based hybrid beamforming for multiuser mmWave MIMO system," Sci. China Inf. Sci., vol. 60, no. 8, 2017.