Research article

Performance analysis of insertion loss incorporated hybrid precoding for massive MIMO

Tadele A. Abose[1],*, Thomas O. Olwal[2], Muna M. Mohammed[3] and Murad R. Hassen[4]

[1] Department of Electrical and Computer Engineering, Mattu University, Mattu, 318, Ethiopia
[2] Department of Electrical Engineering/F'SATI, Tshwane University of Technology, Pretoria, 0001, South Africa
[3] School of Electrical and Computer Engineering, Dire Dawa University, Dire Dawa, 1362, Ethiopia
[4] School of Electrical and Computer Engineering, Addis Ababa University, Addis Ababa, 1176, Ethiopia

* Correspondence: Email: tadenegn@gmail.com; Tel: +251478415516; Fax: +251471414748.

Abstract: Due to an increase in the number of users and a high demand for high data rates, researchers have resorted to boosting the capacity and spectral efficiency of the next-generation wireless communication. With a limited RF chain, hybrid analog digital precoding is an appealing alternative. The hybrid precoding approach divides the beamforming process into an analog beamforming network and a digital beamforming network of a reduced size. As a result, numerous hybrid beamforming networks have been proposed. The practical effects of signal processing in the RF domain, such as the additional power loss incurred by an analog beamforming network, were not taken into account. The effectiveness of hybrid precoding structures for massive MIMO systems was examined in this study. In particular, a viable hardware network realization with insertion loss was developed. Investigating the spectral and energy efficiency of two popular hybrid precoding structures, the fully connected structure, and the subconnected structure, it was found that in a massive MIMO, the subconnected structure always performed better than the fully connected structure. Characterizing the effect of quantized analog precoding, it was shown that the subconnected structure was able to achieve better performance with fewer feedback bits than the fully connected structure.

Keywords: energy efficiency; hybrid precoding; insertion loss; massive MIMO; mm wave

1. Introduction

Due to an increase in the number of users and a high demand for high data rates, researchers have been obligated to boost the capacity and spectral efficiency of next generation wire-less communications. Massive multiple input multiple output (Massive MIMO) technologies, small cell communications, and millimeter wave communications have all been suggested as solutions to boost 5G capacity [1]. To increase 5G capacity, millimeter wave communication takes advantage of a large frequency range between 30 and 300 GHz [2]. Due to its many potential benefits, massive MIMO technology, which makes use of many antennas, has attracted a lot of attention [3–6]. When used individually, each of these technologies can considerably boost the capacity of wireless communication systems, and when utilized collectively, the system's capacity can be further increased [1]. These technologies do, however, face some challenges that must be overcome before they can be effectively implemented. The radio frequency (RF) chains are responsible for a significant portion of the power consumption, hardware cost, and implementation complexity in massive MIMO. They include up converters, digital-to-analog converters, mixers, power amplifiers, and other components [7]. Hybrid analog and digital precoding is a desirable approach because of the constrained RF chains and the possibility to move some signal processing activities to the analog domain [5,6], which, in mm wave [7,8], is significant. Two common hybrid precoding configurations exist: the subconnected structure and the fully connected structure. Performance and complexity trade-offs for subconnected [9,10] and fully connected [7,11] systems were examined.

The subconnected structure is obtained at the expense of somewhat lower performance in order to reduce the complexity, because the fully connected structure has a high level of hardware complexity. Because it reduces performance, the hardware effect must be taken into account in practical implementation. In particular, the insertion loss brought on by the power divider and combiner should not be disregarded because it has a significant impact on the signal power. Various analog beamforming networks have been proposed. However, they frequently ignore the practical effects of RF signal processing, such as the additional power losses brought on by the analog beamforming network (ABFN). Additionally, it contributes to the deterioration of system performance brought on by nonideal hardware, which has a significant impact on a hybrid analog digital processing. The insertion loss brought about by the power divider and combiner, in particular, has a significant impact on the signal power, and should not be disregarded.

Our goal of this research is to analyze the spectral and energy efficiency performance of subconnected and fully connected hybrid precoding structures in a realistic analog processing network, and to enhance performance by identifying an efficient code book.

The following is a summary of this paper's significant contributions: (1) By taking into account insertion loss, which has not been effectively taken into account in the previous literature, we analyzed and compared the spectral and energy efficiency performance of hybrid precoding schemes for subconnected and fully connected archi-tectures. (2) Examining how the energy and spectral efficiency of subconnected and fully connected structures are affected by the phase shifter quantization bit effect when there are many transmit and receive antennas, which is important for mm wave systems.

The rest of the paper has been arranged as follows: The literature review is presented in Section 2, followed by descriptions of the system and channel models, including the Rayleigh fading channel model, the analog precoding network with insertion loss, the quantized version of hybrid precoding,

and the power consumption model. The simulation results are presented in Section 4, and the paper is concluded in Section 5.

2. Literature review

In this section, we aim to review some of the material that is pertinent to the effectiveness of hybrid precoding in MU-MIMO systems under the mm wave scenario. We offer a thorough analysis of hybrid precoding, which has been studied for a decade from a variety of angles. The two types of hybrid precoding—fully connected and sub-connected—are briefly discussed before we begin. The following section looks at hybrid precoding from an ideal to a practical standpoint and models their radio frequency (RF) losses. We then go over a review paper's synopsis on concerns with hybrid precoding that are pertinent to our study.

The performance of the multiuser multiple-input single-output (MU-MISO) system's fully connected and subconnected design was examined by the authors in [12]. They used the same total transmission power constraint to derive the closed-form sum rate of the two designs. According to the findings, fully linked architecture always per-forms better than subconnected architecture in terms of system performance. Insertion loss has not been taken into account for performance analysis, though. MU-MISO, but not massive MIMO, was considered.

Using a phase shifter and a gain controller to drive each antenna element and control the phase of the transmitted signal at each antenna element before or after RF up conversion, analog beamforming techniques are seen as a solution in [13]. The insertion and other losses, however, have gone unnoticed.

To further increase the spectrum efficiency, the authors in [14] studied a mm wave multiuser system. To create a two-stage hybrid beamforming technique, millimeter wave channels with correlated estimate errors and subconnected structures are taken into consideration. The analog components of beam-formers are created to maximize RF-to-RF channel gains in the first stage. The digital components of the beam-formers are optimized in the second step using the equivalent of the maximization of mutual information. They demonstrated that the suggested approach performs better than other existing designs. Insertion loss, however, has not been taken into account for performance evaluations.

The sum rate maximization problem in the subconnected architecture was investigated by the authors of [15]. In order to obtain a relaxed upper limit for the original problem, they first relax the objective function. Then they suggested an approach to finding a local optimal solution to the initial problem. They demonstrated that the suggested local optimal algorithm performs better than the baseline methods in terms of total rate and energy efficiency. However, insertion loss has not been considered for performance analysis.

According to the author in [16], in order to improve performance, a special hard-ware design with a constrained number of fixed-phase shifters was created, and this hardware was then complemented with a dynamic switch network that is channel-adaptive. The suggested fixed phase shifters (FPS) fully connected structure can get close to the performance of a completely digital precoder with a minimal amount of RF chains and phase shifters. A new hybrid precoding mapping method was also presented. Compared to previous mapping techniques that provide two extreme cases, namely fully and partially connected mapping strategies, the proposed group-connected mapping method offers more refined trade-offs between hardware complexity and spectral efficiency.

More importantly, the hybrid precoder can be effectively created utilizing the current hybrid precoding approaches, and this new mapping is compatible with a range of hardware implementations.

The hybrid (analog/digital) beamforming architecture of a multiuser mm wave massive MIMO system with a subconnected configuration was examined by the authors in [17]. A two-stage design approach is used in consideration of the joint receiver and transmitter designs. They demonstrated that the proposed scheme performs better than the most advanced MIMO hybrid beamforming design approaches, and is closer to the performance of a full-digital system. Insertion loss has not been taken into account for performance analysis yet.

In order to realistically assess the benefits of implementing hybrid analog-digital precoding systems in 5G systems, the author of [18] revisits these systems with a focus on the modelling of their radiofrequency (RF) losses. Additionally, they concentrate on discrete Fourier transform implementations and fully connected analog beamforming networks (FC-ABFNs), and they break these down into a bank of widely used RF components. Then, using their S-parameters, they model their losses. The results demonstrate that realistic hybrid scheme performance is (1) highly dependent on hardware implementation, and (2) significantly reduced when realistic losses are taken into ac-count. Only the fully connected structure with discreet fourier transform (DFT) has been taken into consideration in this study. Only the SE metric has been taken into account in the performance study.

In the work of [19], the author contrasts various precoding techniques and offers a straightforward RF system model for the precoder, along with information on how they affect the rate that can be attained. Additionally, they demonstrate how the hybrid precoding system deteriorates significantly once the constraints of the RF precoder network are taken into account. Their results demonstrated that realistic microwave im-plementations significantly degrade the performance of practical hybrid precoding methods. They also noted that the performance degrades as the quantization error associated with F_RF representation using realistic microwave elements increases. On the one hand, they suggest that the RFPNs need to be redesigned by taking RF effects and constraints into account. On the other hand, they offer pointers that RFPNs have made to resemble DFT matrices for large-scale arrays.

The authors of [20] investigate the performance of the fully connected (FC) and one-stream-per-subarray (OSPS) architectures for hybrid digital analog MU-MIMO. In the former, each RF antenna port is connected to each antenna element of the array, while in the latter, the RF antenna ports are connected to separate subarrays. They take into account both the initial beam acquisition phase and the data communication phase, with the latter employing the knowledge of the beam direction acquired in the former. They offer unique BA and precoding strategies for each phase that outperform their literary counterparts. Additionally, they assess the two HDA architectures' power efficiency, while taking into account practical hardware impairments such as power loss at various hardware parts, and probable power back off due to conventional power amplifier (PA) limitations.

According to researchers in [21], quantized hybrid transmitters with fully or partially connected phase-shifting networks composed of active or passive phase-shifters are more energy efficient than quantized digital precoders. They present a quantized single-user MIMO system model based on an additive quantization noise approximation with realistic power consumption and loss models in order to assess the spectrum and energy efficiency of the transmit precoding methodologies. The

simulation findings showed that while fully linked hybrid precoders are typically inefficient, partially connected hybrid precoders are more energy efficient than digital precoders. A balance between energy and spectral efficiency is also provided by the architecture of phase-shifting components: active phase-shifters provide higher data rates, while passive phase-shifters maintain superior energy efficiency. The DAC and quantization effects have been taken into account in this study. However, the power divider and combiner, which cause considerable loss, have not been taken into account. Only the SE metric has been taken into account in the performance study.

The mm wave MIMO system proposed in [22] generalizes the state-of-the-art by taking the inescapable residual transceiver hardware impairments (RTHIs) into account. They provided insight into the effects of three significant hardware impairments: residual additive transceiver hardware impairments (RATHIs), amplified thermal noise (ATN), and multiplicative phase noise (PN). They specifically calculate the system's spectral efficiency in the presence of RTHIs, and offer a thorough study to quantify the damage each individual impairment does to the system's spectral efficiency. The power divider and power combiner, which cause signal power reduction, were not taken into account in this study.

The authors of [23] suggested using linear precoding and effective user and antenna selection algorithms to maximize the system sum-rate capacity. The effectiveness of three user scheduling methods with a joint antenna and user selection has been investigated in order to reduce inter-user interference and improve ergodic sum-rate capacity by linear precoding. In order to increase the system sum and lower the complexity, the authors in [24] suggested a combined semi-orthogonal antenna selection and user selection approach based on a precoding scheme. In order to investigate the system sum-rate performance of a massive MIMO system, the authors in [25] suggested a joint user and antenna selection algorithm where users are scheduled using semi-orthogonality measures and antenna selection is based on maximum channel gain.

The aforesaid literature shows that there are various works on hybrid beamforming that have been done considering different performance metrics, including spectral efficiency, energy efficiency, and system complexity. However, the works that considered the practical effects of signal processing in the RF domain, such as the additional power loss incurred by an analog beamforming network considering the aforementioned performance metrics, were not taken into account. Hence, our aim is to analyze the spectral and energy efficiency performance of hybrid beamforming for massive MIMO systems that takes into account insertion loss.

3. System and channel model

The performance of subconnected and fully connected hybrid analog digital pre-coding with insertion loss is taken into consideration and analyzed in this section. Performance is demonstrated using spectral efficiency and energy efficiency metrics by varying parameters for the massive MIMO scenario.

3.1. System model

Using hybrid analog digital precoding, we focus on the downlink of multiuser massive MIMO in this part. The system consists of a base station (BS) with M antennas, K users with single antennas, and K RF chains, where K is much fewer than M.
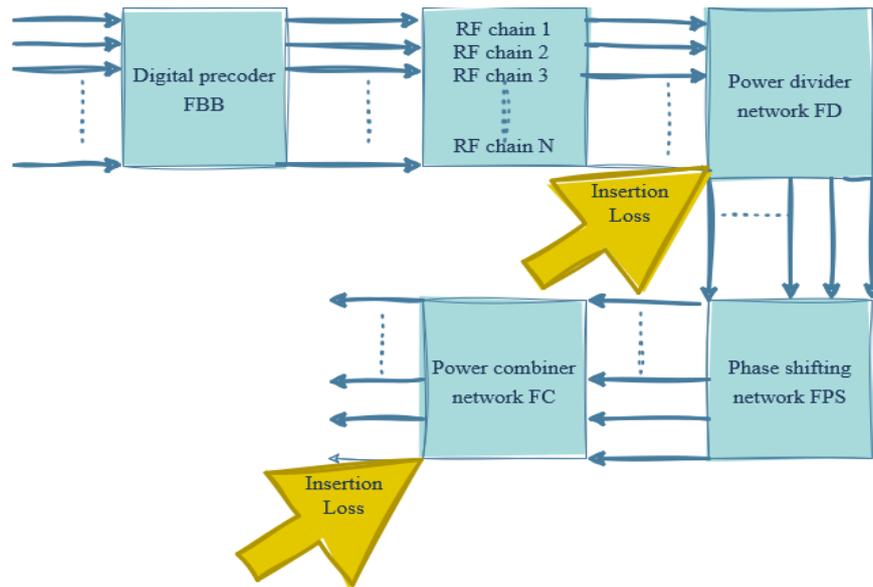
Figure 1. Precoding in a massive MIMO system with M antennas at base station communicating with N users.

Each RF chain at BS is connected to a particular antenna by a phase shifter. Generally speaking, subconnected structures and fully connected structures are the two most common configurations for phase shifter networks connecting RF chains with antennas. In the fully connected structure, each RF chain drives every antenna, and the signals from all RF chains are combined before being fed to the antenna. In the subconnected structure, each RF chain is connected to a disjoint subset of antennas.

The received signal for all K users can be represented as:

$$y = H^H F_{RF} F_{BB} s + n \tag{1}$$

where $S \epsilon C^{k*1}$ represents the transmit symbol vector with $E\{ss^H\} = \frac{p}{k} I_k$, in which p is the transmit power at *BS*. n denotes the additive white Gaussian noise vector with $n \sim C \mathcal{N} (0_k, I_k)$. $H = [h_1, h_2, \ldots \ldots, h_k] \, \varepsilon \, C^{M*K}$ denotes the channel matrix between *BS* and all users with $h_k \sim C \mathcal{N} (0_k, I_k)$, $F_{RF} = [F_{RF,1}, F_{RF,2}, \ldots \ldots \ldots, F_{RF,k}] \, \varepsilon \, C^{M*K}$ represents the analog precoding matrix, and $F_{BB} = [F_{BB,1}, F_{BB,2}, \ldots \ldots \ldots \ldots, F_{BB,k}] \varepsilon \, C^{K*K}$ is the digital precoding matrix.

3.2. Rayleigh fading channel model

The model assumes that the sum of a set of statistically distinct reflected and scattered paths with random amplitudes is an independent, identically distributed (i.i.d.) complex Gaussian random variable. In complex number notation, the channel matrix element can be written as [26]:

$$h_k^m = c + jd \tag{2}$$

Expressing h_k^m in polar form we obtain

$$h_k^m = re^{j\theta} \tag{3}$$

where $r = c^2 + d^2, \theta = \arctan\left(\frac{d}{c}\right)$.

The probability density function (PDF) of Rayleigh fading (R) can be written as.

$$PR(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \qquad (4)$$

The amplitude and phase of a Rayleigh fading channel can be characterized in terms of the aforementioned attributes.

The real and imaginary components of each Gaussian random variable are filled in to generate the complex Rayleigh fading channel coefficient, which is then calculated as follows:

$$h_{Rayleig\ h} = c + jd \qquad (5)$$

where $c \in \mathcal{N}(0, \sigma)$ and $d \in \mathcal{N}(0, \sigma)$.

3.3. Analog precoding network with insertion loss

The signal from each RF chain is first divided into numerous outputs of equal power, as seen in Figure 1. In the subconnected structure, the phase-shifted signal is supplied directly, but in the fully connected design (transmission), the signal is mixed before being fed to the antenna.

It was found that the phase shifter, combiner, and power divider networks—all of which are fully interconnected—make up the analog precoding network. The subconnected structure, however, only consists of divider and phase shifter networks. In order to describe the impact of power loss brought on by hardware networks, it is necessary to separately describe different networks based on the S-parameter of the hardware components.

In a subconnected structure, the analog precoding matrix can be represented by [27]:

$$F_{RF} = F_{ps} F_D \qquad (6)$$

where $F_{ps} \in C^{M*M}$ denotes the matrix of the phase shifter network, and assume the system utilizes the popular Wilkinson power divider and combiner, where the insertion loss is related to the number of ports [28]. F_D stands for the dividing network's effect. From this, the matrix of the divider network can be expressed by:

$$F_D = \sqrt{\frac{1}{N}} \begin{bmatrix} 1_N & 0_N & 0_N \\ 0_N & 1_N & 0_N \\ 0_N & 0_N & 1_N \end{bmatrix} \qquad (7)$$

Similarly, the matrix of phase for the shifter network can be expressed by:

$$F_{ps} = \begin{bmatrix} diag(f_1^{ps}) & 0_N & 0_N \\ 0_N & diag(f_2^{ps}) & 0_N \\ 0_N & 0_N & diag(f_k^{ps}) \end{bmatrix} \qquad (8)$$

where $f_k^{ps} \in C^{N*1}$ represents the phase shifters connected to the K-th RF chain. From this, by substituting Equations (7) and (8) into (6), we analyze the analog precoding in subconnected structure as follows:

$$F_{RF} = \sqrt{\frac{1}{N}} \begin{bmatrix} f_1^{ps} & 0_N & 0_N \\ 0_N & f_2^{ps} & 0_N \\ 0_N & 0_N & f_k^{ps} \end{bmatrix} \tag{9}$$

On the other side, in a fully connected structure, the effect of the combiner net-work should also be considered. Which is:

$$F_{RF} = F_D F_{ps} F_C \tag{10}$$

where F_C is stand for the combiner network so, from this, the *RF* chain for the fully connected structure becomes:

$$F_{RF} = \sqrt{\frac{1}{MK}} \left[f_1^{ps}, f_2^{ps}, \dots\dots\dots, f_k^{ps} \right] \tag{11}$$

where $f_k^{ps} \in C^{M*1}$ has a different dimension from that in the subconnected structure.

3.4. Quantized version of hybrid precoding

In this section, a quantized version of hybrid precoding will be presented. Since the majority of phase shifters in contemporary communication systems are digitally controlled, when designing analog precoding using phase shifters, the angle element is typically quantized and chosen from a set-size codebook [7,29,30]. As a result, the angle of each phase shifter is determined using the least Euclidean distance criterion and a codebook $\mathcal{A} = \left\{ e^{\frac{j2\pi n}{2^B}}, n = 0,1,2,3, \dots\dots, 2^B - 1 \right\}$, where B is the quantity of quantization bits for each phase shifter's analog precoding.

The phase shifter network in the subconnected structure, the *j*-th element of $\widetilde{f_k^{ps}}$, i.e., a quantized version of f_k^{ps}, is normalized by:

$$\widetilde{f_{k,j}^{ps}} = \frac{1}{\sqrt{N}} e^{j \widetilde{\varphi} k,j} \tag{12}$$

where $\tilde{\varphi}_{k,j}$ is a quantized angle satisfying $\tilde{\varphi}_{k,j} \in F_{RF}$. From Equations (9) and (11), the *I*,*j*-th element of $\tilde{F}_{RF}$, a quantized version of F_{RF}, can be written by:

$$\tilde{f}_{k,j} = \begin{cases} \frac{1}{N} e^{j \widetilde{\varphi} k,j} , N(K-1) + 1 \leq j \leq NK \\ 0 , otherwise \end{cases} \tag{13}$$

The same is true for the phase shifter in the fully connected structure $\tilde{f}_{k,j}$:

$$\tilde{f}_{k,j} = \frac{1}{M\sqrt{k}} e^{j \widetilde{\varphi} k,j} \tag{14}$$

In order to maximize the signal strength for each user, the best angles are selected to construct the analog precoding. In this step, user interference is not taken into account, as the analog precoding boosts the received signal power as much as feasible. In order to implement the analog network, $\tilde{f}_{k,j}$ is chosen:

$$\tilde{f}_{k,j} = arg \max_{\tilde{f}_{k,j} \in F_{RF}} \left| h^*_{k,j} \, \tilde{f}_{k,j} \right| \tag{15}$$

where $h_{k,j}$ is the j-th element of h_k.

Digital precoding uses the ZF criterion (c = $H^H (HH^H)^{-1}$) to reduce user interference. In the high SNR regime, the ZF precoding technique, which is provided by pseudo-inverse of the channel matrix, is capable of eliminating interference from other users in the cell. However, its performance in a noise-limited environment is far from ideal. It is important to note that the effective channel affects the ZF precoding for both structures, and that they are similar in several aspects. For user K, an effective channel is defined as:

$$g_k^H \triangleq h_k^H \tilde{F} \tag{16}$$

Each user K quantizes its effective channel vector using a codebook $\mathcal{G}$ of size 2^B according to:

$$\tilde{g}_k = arg \max_{\tilde{g}_k \in \mathcal{G}} |g_k^H \tilde{g}_k| \tag{17}$$

which gives the quantization version of g_k. Each user feeds back the quantized effective channel vector, $\tilde{g}_k$, with B bits.

$$\tilde{F}_{BB} = \tilde{G}(\tilde{G}^H \tilde{G})^{-1} \tag{18}$$

where $\tilde{G} = [\tilde{g}_1, \tilde{g}_2, \dots \dots \dots, \tilde{g}_k]$ and $\tilde{F}_{BB} = [\tilde{f}_1, \tilde{f}_2, \dots \dots \dots \dots, \tilde{f}_k]$ the digital precoder is finally normalized as $\tilde{F}_{BB} = \frac{\tilde{F}_{BB\ k}}{||\tilde{F}_{RF}\tilde{F}_{BB\ k}||}$ to fulfill the power constraints, which is:

$$\left|\left|\tilde{F}_{RF}\tilde{F}_{BB\ k}\right|\right| = 1, K = 1,2,3, \dots \dots \dots \dots \dots, K$$

3.5. The effect of quantized analog precoding

In this section, the effect of quantization on analog and digital precoding will be addressed. Owing to the combination of analog and digital precoding techniques in hybrid precoding, the consequences of quantization for analog and digital precoding should be discussed individually.

Assume perfect digital precoding quantization in this subsection, and because of the nature of ZF, there is no user interference. From Equations (1), (14), and (17), the received signal of the K-th user is given by:

$$y_k = h_k^H \tilde{F}_{RF} F_{BB\ k} s_k + n_k \tag{19}$$

Then, the achievable rates of K-th user and spectral efficiency (*SE*) can be denoted by:

$$R_k = B.\log_2\left(1 + \frac{P}{K}\left|h_k^H \tilde{F} W_k\right|^2\right) = B.\log_2\left(1 + \frac{p}{k}|g_k^H W_k|^2\right) \tag{20}$$

$$SE_k = \log_2\left(1 + \frac{P}{K}\left|h_k^H \tilde{F} W_k\right|^2\right) = \log_2\left(1 + \frac{p}{k}|g_k^H W_k|^2\right) \tag{21}$$

Since $R_k = \log_2(1 + SINR)$, where $SINR = \frac{desired\ signal\ ^2}{Interference\ ^2 + noise\ ^2}$, $B = $ bandwidth. SINR could be written in terms of desired signal power divided by the sum of interfering signal power and noise signal power. The term square could be used since power is proportional to the square of the signals.

Therefore, since a digital precoder is relevant to them, first look into the effective channel vectors. The effective channel in hybrid analog and digital huge MIMO systems is characterized as:

$$g_{k,k} \xrightarrow{a.s} sinc\left(\frac{\pi}{2^B}\right)\frac{\sqrt{\pi}}{2} \tag{22}$$

For the subconnected structure:

$$g_{k,k} \xrightarrow{a.s} sinc\left(\frac{\pi}{2^B}\right)\frac{1}{2}\sqrt{\frac{\pi}{K}} \tag{23}$$

For the fully connected structure, the proof is in Appendix A. According to Equations (20) and (22),

$$G \xrightarrow{a.s} \rho I \tag{24}$$

in which, $\rho = sinc\left(\frac{\pi}{2^B}\right)\frac{\sqrt{\pi}}{2}$ is for the subconnected structure and $\rho = sinc\left(\frac{\pi}{2^B}\right)\frac{1}{2}\sqrt{\frac{\pi}{K}}$ is for the fully connected structure. Therefore, the digital precoder can be expressed by $F_{BB} \xrightarrow{a.s} I$ according to Equation (17), which implies:

$$g_k^H F_{BB,k} \xrightarrow{a.s} g_{k,k} \tag{25}$$

From this, combining Equations (19) and (24), it can be:

$$R_k \xrightarrow{a.s} B.\log_2\left(1 + \frac{P}{K}|g_{k,K}|^2\right) \tag{26}$$

We obtain the results of the achievable rates by substituting Equations (20) and (22) into Equation (25). Whereas for a massive MIMO system, the per-user hybrid precoding attainable rates and SE for a subconnected and a fully connected structure can be roughly represented by:

For the subconnected structure:

$$R_k \xrightarrow{a.s} B.\log_2\left(1 + sinc^2\left(\frac{\pi}{2^B}\right)\frac{\pi P}{4K}\right) \tag{27}$$

For the fully connected structure:

$$R_k \xrightarrow{a.s} B.\log_2\left(1 + sinc^2\left(\frac{\pi}{2^B}\right)\frac{\pi P}{4K^2}\right) \tag{28}$$

For the subconnected structure:

$$SE_k \xrightarrow{a.s} \log_2\left(1 + sinc^2\left(\frac{\pi}{2^B}\right)\frac{\pi P}{4K}\right) \tag{29}$$

For the fully connected structure:

$$SE_k \xrightarrow{a.s} \log_2 \left(1 + sinc^2 \left(\frac{\pi}{2^B} \right) \frac{\pi P}{4K^2} \right) \tag{30}$$

Note: Because a fully connected phase shifter network can achieve more precise analog beamforming, the fully connected structure was once thought to have greater performance. However, the advantages in SINR obtained by accurate analog beamforming are negated by the higher insertion loss in the fully connected structure brought on by the divider network when signals are divided into multiple streams. Additionally, the SINR is further affected by the additional combiner network. The SINR for a subconnected structure is demonstrated to be K times larger than that for a fully connected structure. Therefore, in the case of a large antenna array, the sublinked structure is able to not only attain a greater rate, but also benefit from less hardware complexity [31].

3.6. Some explanation about K

In this section, the continuous positive variable K utilized in analog and digital precoding will be explained. Some observations on the number of RF chains have been made as an indication of the achievable rate. First, look at the subconnected structure's possible rate. According to the research, the achievable rate for perfect digital quantization is [32]:

$$f(k) \triangleq kR_k \tag{31}$$

For the sake of analysis, temporarily relax K as a continuous positive variable. Then, we are able to check the derivation of $f(k)$ with respect to K. It gives

$$f'(k) = \log_2 \left(1 + \frac{\xi}{K} \right) - \frac{\xi}{(K+\xi)ln2} \text{ and } f''(k) = \frac{\xi}{(K+\xi)ln2} \left(\frac{1}{k+\xi} - \frac{1}{K} \right)$$

where $\xi = \frac{\pi p}{4} sinc^2 \left(\frac{\pi}{2^B} \right)$, it is hard to check

$$f''(k) < 0, \forall k > 1 \tag{32}$$

From Equation (28), we know that $f'(k)$ is a monotonic function for $K > 1$. In order to determine the range of $f'(k)$, we need the following result:

$$\lim_{k \to +\infty} f'(k) = \lim_{k \to +\infty} \log_2 \left(1 + \frac{\xi}{k} \right) - \frac{\xi}{(k+\xi)ln2}$$

$$= \frac{1}{ln2} \lim_{k \to +\infty} \left[\ln \left(1 + \frac{\xi}{k} \right) - \frac{\xi}{k+\xi} \right]$$

$$\xrightarrow{(a)} \frac{1}{ln2} \lim_{k \to +\infty} \left[\frac{\xi}{k} + o \left(\frac{\xi}{k} \right) - \frac{\xi}{k+\xi} \right]$$

$$\approx \frac{1}{ln2} \lim_{k \to +\infty} \left[\frac{\xi}{k} - \frac{\xi}{k+\xi} \right], > 0 \tag{33}$$

where (a) utilizes the Taylor's expansion for $\ln \left(1 + \frac{\xi}{k} \right)$. Combining Equations (31) and (32), we know that $f'(k) > 0$ is always true when $K > 1$, which indicate that $f(k)$ is a monotonic increasing function for $k > 1$.

3.7. Power consumption model

In this section, a power consumption model for fully and partially connected structures will be presented. For massive MIMO systems using fully and partially connected hybrid architectures in the mm wave spectrum, a power consumption model has been developed in this subsection. In order to compensate for path losses with directional transmissions, beamforming is used in mm wave large-array MIMO systems.

In this section, we compare the power consumption of the two hybrid architectures for various values of the array size, the number of RF links, and the quantization bit. Since it is challenging to calculate the dissipated power precisely in general [33–35], we estimate each hardware component's power consumption, and provide solid justifications for our decision. Only the transmitter side is the subject of our analysis.

The following general factors and presumptions are mentioned in [35–37] for the power consumption model: (1) For hybrid precoder transmitters that are both fully and partially connected, the same types of devices are taken into consideration; (2) because RF chains are expensive and have a high power consumption, only a few of them are used for each hybrid precoder technique, and they are connected to the large antenna array via a network of phase shifters; and (3) the power consumption of the proposed system is modelled as the sum of transmit and circuit power consumption (signal processing and fixed system power are neglected).

In this paper, the total power consumption becomes:

$$P_T = P_t + P_c \tag{34}$$

where P_T denotes the total power consumption, $P_t = \left\| P_{BB} P_{RF} \right\|_F^2$ is signal transmission power consumption, and P_c is the circuit power consumptions.

Table 1 shows the values of power consumption of prototype power for the power amplifiers P_{PA}, power for the phase shifters P_{PS}, power for the RF chains P_{RF}, power for the digital-to-analog converters P_{DAC}, and power for the baseband processing P_{BB}.

Table 1. Power consumption model for mm wave hardware.

Hardware Components	Power Consumption in Mw
Power for the power amplifier	20 mW
Power for the phase shifter	30 mW
Power for the digital-to-analog converters	200 mW
Power for the RF chain	40 mW
Power for the baseband processing	200 mW

There are two types of phase shifters in mm wave systems, and the power consumption depends on the type and the resolution of the quantized phases. We denote P_{PA} as the power for the power amplifiers, and we denote P_{ps} as the power for the phase shifters. We take into account only passive phase shifters in this part of the power consumption model because active phase shifters and passive phase shifters both exist. P_{RF} is the power for the RF chains. The following components are part of the RF chain block that is expected to be fully and partially connected: A mixer, a local oscillator, a low-pass filter, and a base-band amplifier. The RF chain's power usage can be expressed as [36].

$$P_{RF} = P_{mixer} + P_{LO} + P_{LPF} + P_{BBamp} \qquad (35)$$

where P_{mixer} is the power for the mixer, P_{LO} is power for the local oscillator, P_{LPF} is power for the low pass filter, P_{BBamp} is power for the baseband power amplifier, and it is assumed that all of the *RF* streams are transmitted at the same frequency. P_{DAC} denotes the power for the digital-to-analog converters, and P_{BB} is the power for the baseband processing. To be able to obtain a comparison of the power dissipated for fully connected and partially connected systems, we will also assume that the baseband processor consumes the same power as a single *DAC*, that is $P_{BB} = P_{DAC}$.

Based on the structures, Equations (35) and (36) may be used to express the power dissipated by the hybrid precoder designs for fully connected and partially connected systems, respectively [36].

$$P_{full} = A_{BS}(S_{RF} + 1)P_{PA} + A_{BS}S_{RF}P_{PS} + S_{RF}(P_{RF} + P_{DAC}) + P_{BB} \qquad (36)$$

$$P_{sub} = A_{BS}P_{PA} + A_{BS}P_{PS} + S_{RF}(P_{RF} + P_{DAC}) + P_{BB} \qquad (37)$$

where P_{full} and P_{sub} denote the total power consumption for fully connected and subconnected structures, respectively, A_{BS} is the number of antennas at the base station, and S_{RF} is number of *RF* chains at the base station.

The sum rate for partially and fully connected systems in Equations (27) and (28), respectively, has been shown. Similarly, we have presented the overall power consumption for fully and partially connected architectures in Equations (36) and (37), respectively. Hence, the energy efficiency of partially connected structures can be written as:

$$\eta = \frac{R_{Ksub}}{P_{sub}} \ bits/joule \qquad (38)$$

Similarly, the energy efficiency of fully connected structures can be written as:

$$\eta = \frac{R_{Kfull}}{P_{full}} \ bits/joule \qquad (39)$$

Then, the total power consumption is given by:

$$P_T = \left|\left|P_{BB}P_{RF}\right|\right|_F^2 + P_{full}\left(P_{sub}\right) \qquad (40)$$

Finally, the energy efficiency for hybrid beamforming architecture is represented as shown in [37–40], where energy efficiency is defined as the achievable sum rate per unit power consumption.

$$\eta = \frac{R_K}{P_T} \ bits/joule \qquad (41)$$

4. Results and discussion

In this section, a performance analysis of hybrid precoding with insertion loss for the practical realization of massive MIMO systems will be provided. Energy efficiency and spectral efficiency are the performance metrics used to assess both fully connected structures and subconnected structures. Performance is assessed by changing the BS parameters while using MATLAB plotting. Energy efficiency and spectral efficiency have each been discussed separately.

4.1. Simulation setup

The simulation setup shown in Table 2 is used to evaluate the performance of fully connected and subconnected hybrid precoding with insertion loss in terms of spectral efficiency and energy efficiency.

Table 2. Simulation parameters.

Parameters	Value assumption
Number of cells	Single cell
System model	Fully connected and subconnected hybrid precoding
Number of data streams	Equal to the number of *RF* chains
Channel model	Rayleigh fading channel
Channel state information (CSI)	Perfect channel knowledge

4.2. Spectral efficiency evaluation

We consider massive MU-MIMO systems with insertion loss. Spectral efficiency in bps/Hz is the number of bits of information per complex-valued sample that can be reliably transmitted over the mm wave channel under consideration. The results shown under this section are spectral efficiency versus SNR (dB), number of antenna arrays, number of quantization bits, and number of RF chains.

4.2.1. Spectral efficiency with varying SNR

The performance of massive MU-MIMO systems with insertion loss over Rayleigh fading channels is investigated by considering the achievable rate. The number of users at BS is 4, and the number of antennas at BS is 128, with quantization bit B = 1, 3. In Figure 2, the spectral efficiency severely increases with the number of analog quantization bits and transmit power. Moreover, the system performance of the subconnected structure is better than that of the fully connected structure due to the parameter, the continuous positive variable, K, as shown in (29–30). The spectral efficiency is related to 1/K in a subconnected structure (29) and to $1/K^2$ in a fully connected structure (30). As the number of quantization bits increases, the spectral efficiency increases because of the reduction of phase error (noise). In this graph, when the SNR is high (20 dB), the fully connected structure at bit 3 becomes 12 and at bit 1 it is 8, but in the subconnected structure at 20 dB, the subconnected structure at bit 3 becomes 22, and at bit 1 it is 18. Hence, the subconnected structure exhibits better performance than the fully connected structure because of the consideration of hardware losses (insertion loss). In the fully connected structure, high insertion loss introduced by the power divider network, as the signal is divided into more streams cancels the benefit in SINR acquired by accurate analog beamforming, and in the subconnected structure, the achievable rate is k times greater than the fully connected structure.
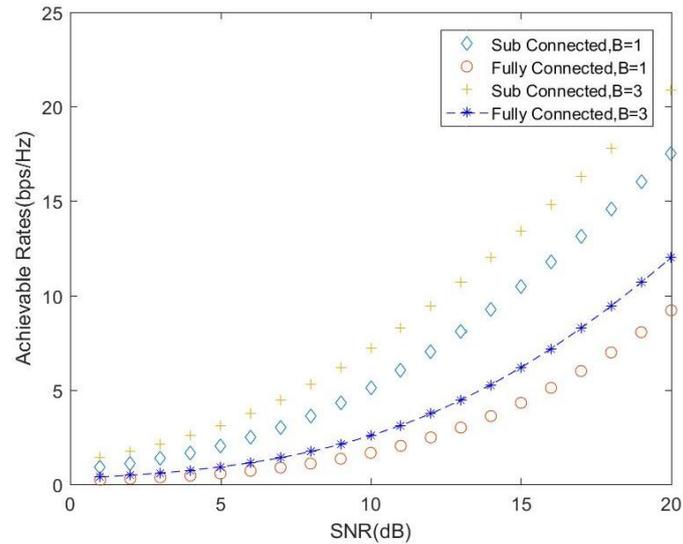
Figure 2. Spectral efficiency vs. SNR (dB) for A_*BS* = 128 and *K* = 4.

4.2.2. Spectral efficiency vs. number of antenna arrays

Figure 3 shows that the spectral efficiency of a subconnected structure outperforms that of a fully connected structure because, as the number of antennas increases, the antenna array also increases, leading to high spectral efficiency. From this graph, it is observed that the subconnected structure with bits 3 and 1 has better performance than the fully connected structure because the subconnected structure is K-times larger than that of the fully connected structure, with the consideration of practical losses (insertion loss), which cause performance degradation in the fully connected structure.
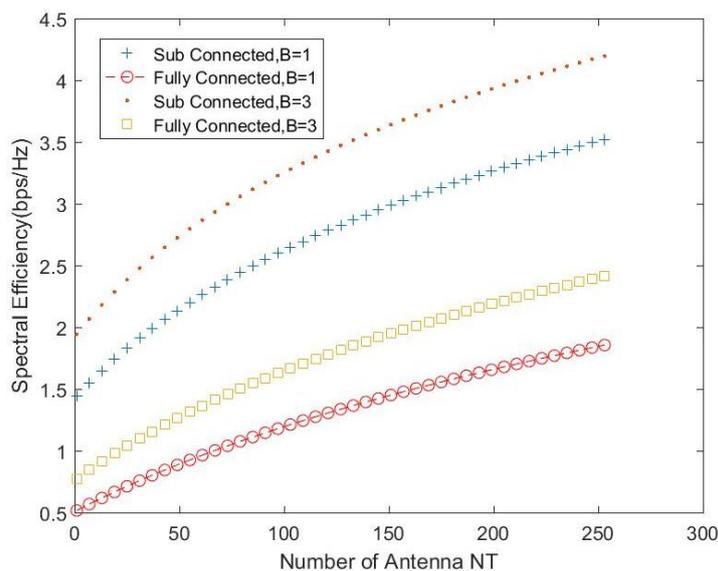
Figure 3. Spectral efficiency vs. number of antenna arrays.

The simulation time, memory requirement, and processing requirement increase with the increasing number of antennas. From a practical limitation perspective, the study decided to run

simulations with the maximum number of antennas set to 250; however, any other higher values could be possible without any impact on the spectral efficiency performance, but that requires more computational time.

4.2.3. Spectral efficiency vs. number of quantization bits

Figure 4 shows the comparison between fully connected structures and subconnected structures in terms of spectral efficiency vs. number of quantization bits. The subconnected structure outperforms the fully connected structure because of hardware loss considerations, as we see from Figure 4. As the number of quantization bits increases, the phase error (noise) decreases because each data stream is quantized. From Figure 4, it can be noticed that when the quantization bit becomes 3, the fully connected structure's spectral efficiency is 1.883, and the subconnected structure's spectral efficiency is 2.089. When the quantization bit becomes 4, the fully connected structure's spectral efficiency is 2.758, and the subconnected structure's spectral efficiency is 2.996. When the quantization bit is 5, the fully connected structure's spectral efficiency is 3.014 and the subconnected structure's spectral efficiency is 3.256. After 5 bits, spectral efficiency goes to saturation or has no difference.
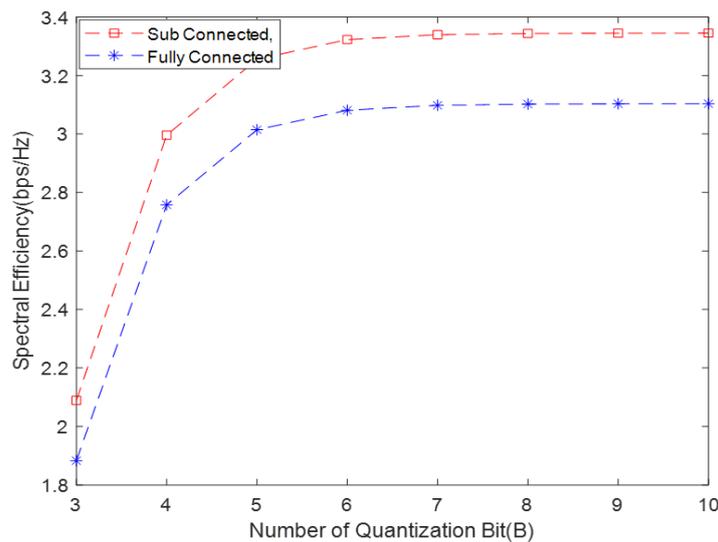
Figure 4. Spectral efficiency vs. number of quantization bits.

4.2.4. Spectral efficiency vs. number of RF chains

Figure 5 shows spectral efficiency vs. number of RF chains, and we see that as the number of RF chains increases, the subconnected structure with bit 3 outperforms the fully connected structure because of the high insertion loss introduced by the power divider network. From Figure 5, it can be noted that when the number of RF chains is 9, the fully connected structure's spectral efficiency at bit 3 becomes 0.4458, and the subconnected structure's spectral efficiency becomes 0.8612. When the RF chain is 23, the fully connected structure's spectral efficiency is 0.938, and the subconnected structure's spectral efficiency is 1.481.

The optimal number of RF chains for achieving perfect spectral efficiency in this study was 50, which provides 2 bps/Hz for sub-connected and 1.5 bps/Hz for fully-connected structures. When the

number of RF chains doubled, which is 100, there was a spectral efficiency improvement of only 0.5 bps/Hz for the sub-connected structure at the cost of 50 extra RF chains, while there was a spectral efficiency improvement of only 0.7 bps/Hz for the fully connected structure.
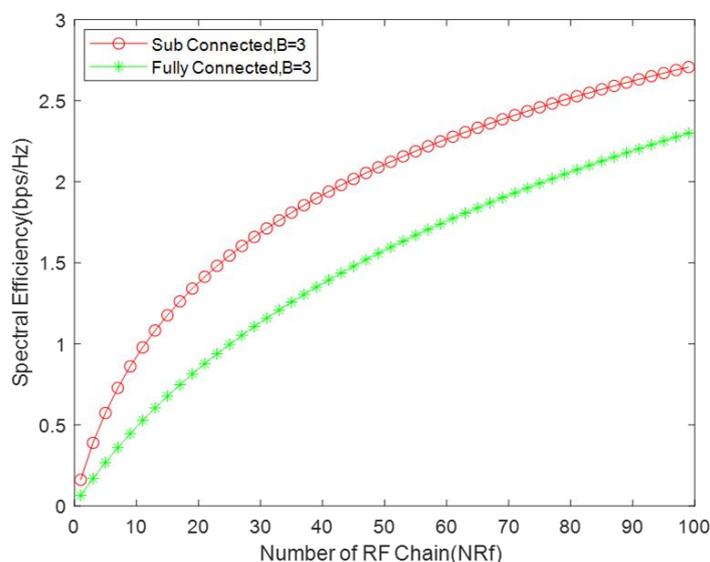
Figure 5. Spectral efficiency vs. number of *RF* chains**.**

4.3. Energy efficiency evaluation

In this subsection, the energy efficiency will be plotted by considering the total power consumption for the signal transmission power and the circuit power consideration. The results shown under this section are energy efficiency versus number of antennas, number of quantization bits, and number of RF chains.

4.3.1.   Energy efficiency vs. number of antenna arrays

Figure 6 shows energy efficiency vs. number of antennas. From Figure 6, it can be seen that, as the number of antennas increases, the energy efficiency decreases or falls. Since energy efficiency is the ratio of data rate to total power consumed, a large number of antennas causes a greater rise in total power consumption than an increase in data rate when insertion loss is taken into account. Therefore, as the number of antennas rises, energy efficiency falls. The fully connected structure's energy efficiency with bits 3 and 1 decreases as compared to that of the subconnected structure because it consumes more energy. As the number of antennas increases, the energy required also increases, so a fully connected structure consumes more energy.
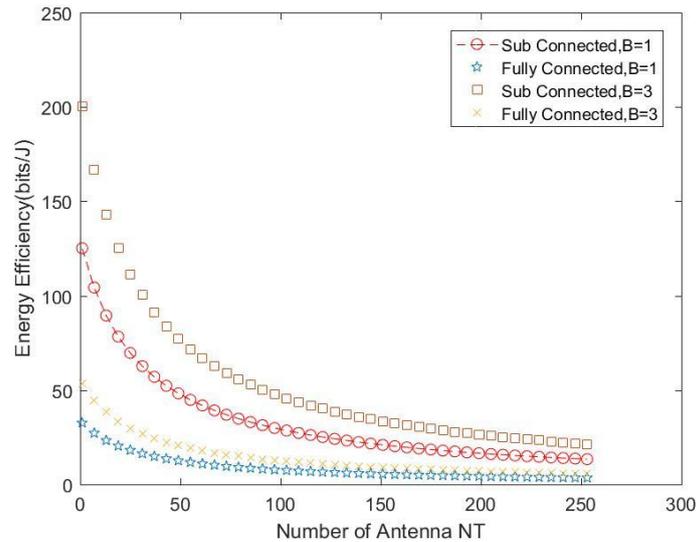
Figure 6. Energy efficiency vs. number of antenna array.

4.3.2. Energy efficiency vs. number of quantization bits

Figure 7 shows energy efficiency vs. quantization bits. From Figure 7, it can be seen that, as energy efficiency decreases, the number of quantization bits increases. When the quantization bit becomes 4, both the fully connected and subconnected structures decrease because the energy consumption is higher. The energy efficiency was increased up to 4 quantization bits for both the fully connected and subconnected structures and it gradually decreased with an increasing number of bits. For quantization bits up to 4, the accuracy of quantization improved, which in turn improved phase shifter precision. The better accuracy results in better beamforming performance for quantization bits of up to 4, which leads to an improvement in energy efficiency. After quantization bits of 4, quantization accuracy improvement starts to decrease. Any addition of quantization bits demands more power, which means energy efficiency starts to decrease.
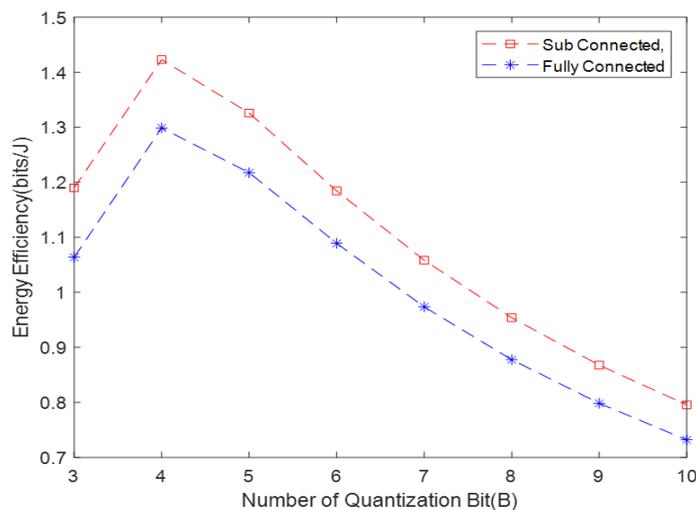
Figure 7. Energy efficiency vs. number of quantization bits.

4.3.3. Energy efficiency vs. number of RF chains

Figure 8 shows energy efficiency vs. number of RF chains. From Figure 8, it can be seen that, as the number of RF chains increases, the energy efficiency decreases because the RF chain consumes more energy.
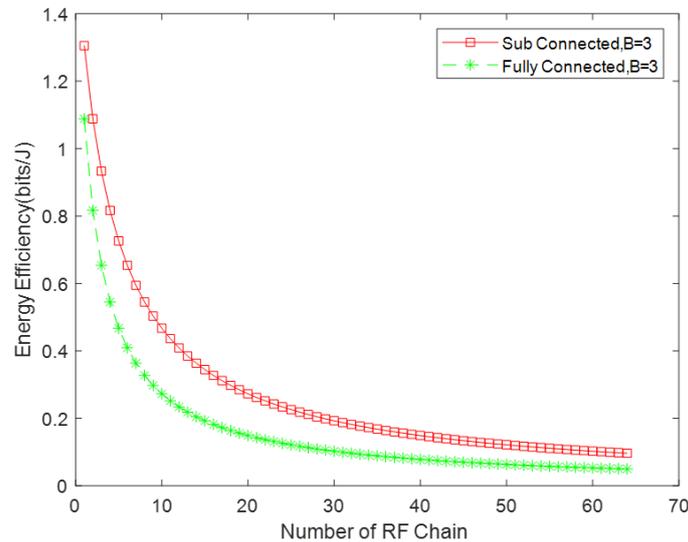
Figure 8. Energy Efficiency vs. number of *RF* chains.

5. Conclusions

We evaluate the performance of fully connected structures and subconnected structures in hybrid precoding with insertion loss. With consideration of insertion loss, the performance is compared for fully connected and subconnected structures using the spectral efficiency metric and the energy efficiency metric. The simulation results show that the system performance, in terms of spectral efficiency, is better in the subconnected structure than in the more complex fully connected structure in a system with a massive antenna array. From the simulation results, as the number of quantization bits increases, when the quantization bit becomes 3, the fully connected SE is 1.883, and the subconnected SE is 2.089, and when the quantization bit becomes 4, the fully connected SE is 2.758 and the subconnected SE is 2.996. When the quantization bit is set to 5, the fully connected SE is 3.014 and the subconnected SE is 3.256. After that, they go to saturation or have no difference. From this, we can observe that the subconnected structure outperforms the fully connected structure. When the number of RF chains is 9, the fully connected SE at bit 3 becomes 0.4458 and the SE of the subconnected becomes 0.8612, and when the high RF chain is 23, the fully connected SE is 0.938 and the subconnected SE is 1.481. From this, we can observe that the subconnected outperform the fully connected, and as the number of RF chains increases, the difference between them becomes bigger. It can also be observed that, as the number of RF chains increases, subconnected structures are more energy efficient than fully connected structures. The future work of this paper will include analyzing the performance of hybrid precoding multiuser and multicell massive-MIMO systems with other deep hardware losses, such as static and distortion losses.

Use of AI tools declaration

The authors declare that they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

The authors declare that there are no conflicts of interest in this paper.

References

1. Swindlehurst AL, Ayanoglu E, Heydari P, Capolino F (2014) Millimeter-wave massive MIMO: The next wireless revolution? *IEEE Commun Mag* 52: 56–62. https://doi.org/10.1109/MCOM.2014.6894453

2. Karjalainen J, Nekovee M, Benn H, Kim W, Park J, Sungsoo H (2014) Challenges and Opportunities of mm-Wave Communication in 5G Networks. *Proceedings of the 2014 9th International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM)*, 372–376. https://doi.org/10.4108/icst.crowncom.2014.255604

3. Marzetta TL (2010) Noncooperative cellular wireless with unlimited numbers of base station antennas. *IEEE T Wirel Commun* 9: 3590–3600. https://doi.org/10.1109/TWC.2010.092810.091092

4. Larsson EG, Edfors O, Tufvesson F, Marzetta TL (2014) Massive MIMO for next generation wireless systems. *IEEE Commun Mag* 52: 186–195. https://doi.org/10.1109/MCOM.2014.6736761

5. Xie H, Wang B, Gao F, Jin S (2016) A full-space spectrum-sharing strategy for massive MIMO cognitive radio systems. *IEEE J Sel Areas Commun* 34: 2537–2549. https://doi.org/10.1109/JSAC.2016.2605238

6. Feng W, Gao F, Shi R, Ge N, Lu J (2015) Dynamic-cell-based macro coordination for massively distributed MIMO systems. *Proceedings of the 2015 IEEE Global Communications Conference (GLOBECOM)*, 1–6. https://doi.org/10.1109/GLOCOM.2015.7417293

7. Ayach OE, Rajagopal S, Abu-Surra S, Pi Z, Heath RW (2014) Spatially sparse precoding in millimeter wave MIMO systems. *IEEE T Wirel Commun* 13: 1499–1513. https://doi.org/10.1109/TWC.2014.011714.130846

8. Roh W, Seol JY, Park J, Lee B, Lee J, Kim Y, et al. (2014) Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results. *IEEE Commun Mag* 52: 106–113. https://doi.org/10.1109/MCOM.2014.6736750

9. Han S, Chih-Lin I, Xu Z, Rowell C (2015) Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G. *IEEE Commun Mag* 53: 186–194. https://doi.org/10.1109/MCOM.2015.7010533

10. Gao X, Dai L, Han S, Chih-Lin I, Heath RW (2016) Energy-efficient hybrid analog and digital precoding for mmwave MIMO systems with large antenna arrays. *IEEE J Sel Areas Commun* 34: 998–1009. https://doi.org/10.1109/JSAC.2016.2549418

11. Liang L, Xu W, Dong X (2014) Low-complexity hybrid precoding in massive multiuser MIMO systems. *IEEE Wireless Commun Lett* 3: 653–656. https://doi.org/10.1109/LWC.2014.2363831

12. Wan S, Zhu H, Kang K, Qian H (2021) On the Performance of Fully-Connected and Sub-Connected Hybrid Beamforming System. *IEEE T Veh Technol* 70: 11078–11082. https://doi.org/10.1109/TVT.2021.3109300

13. Ngo HQ (2015) *Massive MIMO: Fundamentals and System Designs*. Vol. 1642, Linköping University Electronic Press: Linköping, Sweden. https://doi.org/10.3384/lic.diva-112780

14. Luo Z, Luo L, Zhang X, Liu H (2022) Robust Hybrid Beamforming for Multi-user Millimeter Wave Systems with Sub-connected Structure. *International Conference on Communications and Networking in China.* https://doi.org/10.1007/978-3-031-34790-0_9

15. Hu Y, Qian H, Kang K, Luo X, Zhu H (2023) Joint Precoding Design for Sub-Connected Hybrid Beamforming System. *IEEE T Wirel Commun*. https://doi.org/10.1109/TWC.2023.3287229

16. Yu X, Zhang J, Letaief KB (2018) A Hardware-Efficient Analog Network Structure for Hybrid Precoding in Millimeter Wave Systems. *IEEE J Sel Top Signal Process* 12: 282–297. https://doi.org/10.1109/JSTSP.2018.2814009

17. Zhang Y, Du J, Chen Y, Li X, Rabie KM, Khkrel R (2020) Dual-Iterative Hybrid Beamforming Design for Millimeter-Wave Massive Multi-User MIMO Systems With Sub-Connected Structure. *IEEE T Veh Technol* 69: 13482–13496. https://doi.org/10.1109/TVT.2020.3029080

18. Garcia-Rodriguez A, Venkateswaran V, Rulikowski P, Masouros C (2016) Hybrid analog-digital precoding revisited under realistic RF modeling. *IEEE Wireless Commun Lett* 5: 528–531. https://doi.org/10.1109/LWC.2016.2598777

19. Venkateswaran V, Krishnan R (2016) Hybrid Analog and Digital Precoding: From Practical RF System Models to Information Theoretic Bounds. *2016 IEEE Globecom Workshops (GC Wkshps).* https://doi.org/10.1109/GLOCOMW.2016.7848924

20. Song X, Kuhne T, Caire G (2019) Fully-Connected vs. Sub-Connected Hybrid Precoding Architectures for mmWave MU-MIMO. *ICC 2019-2019 IEEE International Conference on Communications (ICC).* https://doi.org/10.1109/ICC.2019.8761521

21. Ribeiro LN, Schwarz S, Rupp M, de Almeida AL (2018) Energy Efficiency of mmWave Massive MIMO Precoding with Low-Resolution DACs. *IEEE Journal of Selected Topics in Signal Processing* 12: 298–312. https://doi.org/10.1109/JSTSP.2018.2824762

22. Kolawole O, Papazafeiropoulos A, Ratnarajah T (2018) Impact of Hardware Impairments on mmWave MIMO Systems with Hybrid Precoding. *Proceedings of the IEEE Wireless Communication and Networking Conference*. https://doi.org/10.1109/WCNC.2018.8377045

23. Sheikh TA, Bora J, Hussain MA (2021) Capacity maximizing in massive MIMO with linear precoding for SSF and LSF channel with perfect CSI. *Digit Commun Netw* 7: 92–99. https://doi.org/10.1016/j.dcan.2019.08.002

24. Sheikh TA, Bora J, Hussain MA (2019) Combined user and antenna selection in massive MIMO using precoding technique. *International Journal of Sensors Wireless Communications and Control* 9: 214–223. https://doi.org/10.2174/2210327908666181112144939

25. Sheikh TA, Bora J, Hussain MA (2018) Sum-rate performance of massive MIMO systems in highly scattering channel with semi-orthogonal and random user selection. *Radioelectronics and Communications Systems* 61: 547–555. https://doi.org/10.3103/S0735272718120026

26. Papoulis A, Unnikrishna Pillai S (2012) *Probability, Random Variables, and Stochastic Processes*, North America: McGraw-Hill, New York, United States.

27. Venkateswaran V, Pivit F, Guan L (2016) Hybrid RF and digital beamformer for cellular networks: Algorithms, microwave architectures, and measurements. *IEEE T Microw Theory Technol* 64: 2226–2243. https://doi.org/10.1109/TMTT.2016.2569583

28. Pozar DM (2009) *Microwave Engineering*, John Wiley & Sons: Hoboken, NJ, USA.

29. Alkhateeb A, Leus G, Heath RW (2015) Limited feedback hybrid precoding for multi-user millimeter wave systems. *IEEE T Wireless Commun* 14: 6481–6494. https://doi.org/10.1109/TWC.2015.2455980

30. Fozooni M, Matthaiou M, Jin S, Alexandropoulos GC (2016) Massive MIMO relaying with hybrid processing. *Proceedings of the 2016 IEEE International Conference on Communications (ICC)*, 1–6. https://doi.org/10.1109/ICC.2016.7510972

31. Du J, Xu W, Shen H, Dongy X, Zhao C (2017) Quantized Hybrid Precoding for Massive Multiuser MIMO with Insertion Loss. *GLOBECOM 2017-2017 IEEE Global Communications Conference*. https://doi.org/10.1109/GLOCOM.2017.8254815

32. Du J, Xu W, Shen H, Dong X, Zhao C (2018) Hybrid Precoding Architecture for Massive Multiuser MIMO with Dissipation: Sub-Connected or Fully-Connected Structures? *IEEE T Wirel Commun* 17: 5465–5479. https://doi.org/10.1109/TWC.2018.2844207

33. Ratnam VV, Molisch AF, Bursalioglu OY, Papadopoulos HC (2018) Hybrid beamforming with selection for multiuser massive mimo systems. *IEEE T Signal Process* 66: 4105–4120. https://doi.org/10.1109/TSP.2018.2838557

34. Wei N (2007) *MIMO Techniques for UTRA Long Term Evolution*, Citeseer: Princeton, NJ, USA.

35. Liu J, Bentley E (2017) Hybrid-beamforming-based millimeter-wave cellular network optimization. *IEEE J Sel Area Commun* 37: 2799–2813. https://doi.org/10.23919/WIOPT.2017.7959916

36. Méndez-Rial R, Rusu C, González-Prelcic N, Alkhateeb A, Heath RW (2016) Hybrid mimo architectures for millimeter wave communications: Phase shifters or switches? *IEEE Access* 4: 247–267. https://doi.org/10.1109/ACCESS.2015.2514261

37. Jedda H, Ayub MM, Munir J, Mezghani A, Nossek JA (2015) Power-and spectral efficient communication system design using 1-bit quantization. *Proceedings of the 2015 International Symposium on Wireless Communication Systems (ISWCS)*, 296–300 https://doi.org/10.1109/ISWCS.2015.7454349

38. Jing J, Xiaoxue C, Yongbin X (2016) Energy-efficiency based downlink multi-user hybrid beamforming for millimeter wave massive mimo system. *J China Univ Posts Telecommun* 23: 53–62. https://doi.org/10.1016/S1005-8885(16)60045-6

39. Zhang Y, Yang Y, Dai L (2016) Energy efficiency maximization for device-to-device communication underlying cellular networks on multiple bands. *IEEE Access* 4: 7682–7691. https://doi.org/10.1109/ACCESS.2016.2623758

40. Abose TA, Olwal TO, Hassen MR, Bekele ES (2022) Performance Analysis and Comparisons of Hybrid Precoding Scheme for Multi-user mmWave Massive MIMO System. *2022 3rd International Conference for Emerging Technology (INCET)*, 1–6. https://doi.org/10.1109/INCET54531.2022.9824401

Appendix A

From Equation (18):

$$g_{k,k} = \left(h_k^H \tilde{f}_k \right)^*$$

$$= \frac{1}{N} \sum_{i=(k-1)N+1}^{KN} h_{k,i} e^{-j\tilde{\varphi}_{k,i}}$$

$$\triangleq \frac{1}{N} \sum_{i=(k-1)N+1}^{KN} \lambda_i \qquad (A1)$$

where $\lambda_i \triangleq h_{k,i} e^{-j\tilde{\varphi}_{k,i}}$.

Define the phase error $\varepsilon_{k,i}$ as the error between the unquantized phase $\varphi_{k,i}$ and quantized phase $\tilde{\varphi}_{k,i}$ i.e., $\varepsilon_{k,i} \triangleq \varphi_{k,i} - \tilde{\varphi}_{k,i}$, which yields:

$$\lambda_i = \left| h_{i,k} \right| e^{j\varepsilon_{k,i}} \qquad (42)$$

Since $e^{j\varepsilon_{k,i}}$ and $\left| h_{i,k} \right|$ are independent, we investigate them separately. As the phase of each entry in channel matrix H follows uniform distribution between 0 and 2π, i.e., $\varphi_{k,j} \sim u[0, 2\pi)$, we can easily derive the distribution of the phase error as $\varepsilon_{k,i} \sim u[-\delta, \delta)$, where we define $\delta \triangleq \frac{\pi}{2^B}$. Applying the Euler's formula, we obtain:

$$\Re\left[e^{j\varepsilon_{k,i}} \right] = cos\varepsilon_{k,i}$$

$$\Im\left[e^{j\varepsilon_{k,i}} \right] = sin\varepsilon_{k,i} \qquad (43)$$

Furthermore, it is easy to obtain:

$$\mathrm{E}\left[\Re\left[e^{j\varepsilon_{k,i}} \right] \right] = \frac{1}{2\delta} \int_{-\delta}^{\delta} cos\varepsilon_{k,i} d\varepsilon_{k,i} = sinc(\delta) \qquad (44)$$

$$\mathrm{E}\left[\Re\left[e^{j\varepsilon_{k,i}} \right] \right] = \frac{1}{2\delta} \int_{-\delta}^{\delta} cos^2\varepsilon_{k,j} d\varepsilon_{k,j} = \frac{1}{2}[1 + sinc(\delta)\cos(\delta)] \qquad (45)$$

where $sinc(\delta) = \frac{\sin(\delta)}{\delta}$.

Recalling $h_k \sim \mathcal{CN}(0_k, I_k)$, $\left| h_{i,k} \right|$ follows Rayleigh distribution, and hence:

$$\mathrm{E}[\left| h_{i,k} \right|] = \frac{\sqrt{\pi}}{2}$$

$$V[\left| h_{i,k} \right|] = 1 - \frac{\pi}{4} \qquad (46)$$

$$\mathrm{E}\left[\left| h_{i,k} \right|^2 \right] = \mathrm{E}^2[\left| h_{i,k} \right|] + v[\left| h_{i,k} \right|] = 1 \qquad (47)$$

Since $\left| h_{i,k} \right|$ is definitely real, it is obvious that

$$\Re[\lambda_i] = \Re\left[\left| h_{i,k} \right| e^{j\varepsilon_{i,k}} \right] = \left| h_{i,k} \right| \Re\left[e^{j\varepsilon_{k,i}} \right]$$

$$\Im[\lambda_i] = \Im\big[|h_{i,k}|e^{j\,\varepsilon_{k,i}}\big] = |h_{i,k}|\Im[e^{j\,\varepsilon_{k,i}}] \tag{48}$$

Owing to the independency between $|h_{i,k}|$ and $e^{j\,\varepsilon_{k,i}}$, we can further obtain that

$$\mathrm{E}\big[\Re[\lambda_i]\big] = \mathrm{E}\big[|h_{i,k}|\big]\mathrm{E}\Big[\Re\big[e^{j\,\varepsilon_{k,i}}\big]\Big] = sinc(\delta)\frac{\sqrt{\pi}}{2} \tag{49}$$

$$\mathrm{E}[(\Re[\lambda_i])^2] = \mathrm{E}\left[\big|h^*{}_{i,k}\big|^2\right]\mathrm{E}\big[(\Re[e^{j\,\varepsilon_{k,i}}])^2\big] = \frac{1}{2}[1 + sinc(\delta)\cos(\delta)] \tag{50}$$

where Equations (A4)–(A7) are used. According to Equations (A9) and (A10), it is apparent that

$$\nu\big[\Re[\lambda_i]\big] = \mathrm{E}[(\Re[\lambda_i])^2] - \big(\mathrm{E}[\Re[\lambda_i]]\big)^2 = w_1 \tag{51}$$

where $w_1 = \frac{1}{2}[1 + sinc(\delta)\cos(\delta) - sinc^2(\delta)\frac{\pi}{4}]$. By applying the central limit theorem and utilizing Equations (A9) and (A10), we obtain:

$$\Re[g_{k,k}] \xrightarrow{a.s} sinc(\delta)\frac{\sqrt{\pi}}{2}$$

$$\Im[g_{k,k}] \xrightarrow{a.s} 0. \tag{52}$$

Following trivially the steps above, it yields:

$$\mathcal{R}[g_{k,j}] \xrightarrow{a.s} 0$$

$$\Im[g_{k,j}] \xrightarrow{a.s} 0 \tag{53}$$

Therefore, from Equations (A12) and (A13), we ultimately achieve Equation (24). The proof for the fully connected structure is similar to that for the subconnected structure.