IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 28, 2022, accepted February 8, 2022, date of publication February 17, 2022, date of current version March 4, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3152537

Hybrid Beamformer Exploiting Multistream per User Transmission for Millimeter-Wave NOMA Communications

ABDULAHI ABIODUN BADRUDEEN[1,2], (Member, IEEE), CHEE YEN LEOW[1], (Senior Member, IEEE), AND SEUNGHWAN WON[3], (Senior Member, IEEE)

[1]Wireless Communication Center, School of Electrical Engineering, Universiti Teknologi Malaysia, Johor Bahru 81310, Malaysia
[2]Department of Computer Engineering Technology, Federal Polytechnic Ede, Ede, Osun State 232104, Nigeria
[3]School of Electronics and Computer Science, University of Southampton Malaysia, Iskandar Puteri 79100, Malaysia

Corresponding author: Chee Yen Leow (bruceleow@utm.my)

This work was supported in part by the Ministry of Higher Education Malaysia; and in part by Universiti Teknologi Malaysia under Grant FRGS/1/2020/TK0/UTM/02/68, Grant A.J091300.6800.07085, Grant 09G15, and Grant 08G83. The work of Abdulahi Abiodun Badrudeen was supported by the Federal Republic of Nigeria through the Tertiary Education Trust Fund (TETFund).

ABSTRACT This treatise investigates multistream hybrid beamforming (HBF) for a millimeter-wave non-orthogonal multiple access (NOMA) system deployed in a downlink of an urban microcell environment. Maximization problem is developed to optimize sum-rate. For the sake of ensuring high correlation of users' channels, user's clustering and ordering are based on the angle of arrivals and users' channel-weights, respectively. An optimized analog combiner of each user is obtained from the first N_s column vectors of the left unitary matrix derived from the singular value decomposition of the channel of each user. Analog beamformer is matched to the phase of the strong users' composite intermediate analog channel to maximize the beamforming gain. Both the analog combiner and precoder for sub-connected structure (SCS) are formulated via a novel dominant sub-array matrix elements' extractor. Conventional zero-forcing processing is contemplated for digital precoding. Memory space complexity is evaluated to corroborate the simplicity of the proposed schemes' computational complexity. Results obtained from the simulation exhibit that HBF-NOMA attain superior sum-rate than HBF-orthogonal multiple access and conventional multiuser schemes in line of sight link. Lastly, the proposed SCS-HBF-NOMA precoding scheme performs higher than fully connected counterpart in terms of energy efficiency.

INDEX TERMS Fully connected structure, millimeter-wave, multistream hybrid beamforming, NOMA, sub-connected structure, sum-rate maximization.

I. INTRODUCTION

Rapid demands for high data-rate mobile wireless applications such as online gaming, telemedicine, electronic learning, and video conferencing result in a bottleneck for deploying below 6 [GHz] spectrum owing to the inherently scarce signal bandwidth limitation. Therefore, a paradigm shift from the current existing spectrum band to a millimeter-wave (mmW) one ranging from 30 [GHz] to 300 [GHz] is inevitable to support high data rate traffic in order of Gbps for the post-5G wireless communications. However, mmW propagation suffers from path loss fading, rain absorption and shadowing in line of sight (LOS). The small wavelength of mmW allowing the exploitation of a large number of antennas in a small area of wireless device circuit is to combat the effect of path loss fading. Deploying massive antennas for multiple input multiple output (MIMO) having fully digital beamfoming (DBF) for future wireless system will lead to namely: 1) high radio frequency (RF) chains' complexity, 2) more power consumption, and 3) high implementation cost. Hence, hybrid (digital and analog) beamforming (HBF) has been proposed as an appropriate beamforming (BF) approach for massive MIMO.

The two major HBF structures named fully connected structure (FCS) and sub-connected structure (SCS) are distinguished exploiting RF chains and antennas' linking mode,

The associate editor coordinating the review of this manuscript and approving it for publication was Ding Xu.

respectively. Specifically, in FCS-HBF, each RF chain is linked to all antennas via a set of phase shifters (PSs) [1], while the SCS-HBF has each RF chain chosen from the aggregate RF chains at the output of the base band precoder mapped to each group of sub-array antennas via the corresponding set of PSs [2], [3]. Consequently, FCS-HBF can attain full BF gain in comparison to SCS-HBF because each of its RF chains is associated with all antennas. The existing works have studied both structures for single user single-carrier and multi-carrier cases in [3]–[8] and [8]–[10], respectively. Moreover, multiuser system has also been studied for large-scale users connectivity [11]–[14]. Specifically, authors in [14] investigated multistream FCS-HBF for conventional multiuser in order to further improve system sum-rates. However, multistream FCS-HBF structure leads to hardware complexity and consume more power than the single stream counterpart.

Moreover, the traditional multiuser's base station (BS) can only serve a maximum number of users that is equal to its RF chains. Non-orthogonal multiple access (NOMA) was suggested in literature to address this limitation of traditional multiuser system. Power domain NOMA has received much attention among researchers [15] owing to its numerous advantages elucidated in [16]. The power domain NOMA leverages superimposition of multiusers' symbols at the BS before transmission and successive interference cancellation (SIC) of the feeble users' symbols by the strong users at the receiver [17]. Mitigation schemes against the intracluster interference and intercluster interference along with obtaining optimal power allocation for multiuser NOMA systems have been a major concern among researchers in a recent decade. Therefore, a proper BF scheme is crucial for removal of intercluster interference and an efficient SIC scheme to eliminate intracluster interference in NOMA systems.

Study of FCS-HBF-NOMA exploiting single stream per user transmission is well known in literature [1] owing to its full BF gain, which can boost transmit signal to the cluster users for effective communications. Similarly, adopting the SCS-HBF to NOMA for single stream per user configuration is firstly known in [18]. Authors carried out their study using simplistic Saleh Valenzuela channel model, which are not realistic because the channel gain is not based on practical measurement such as the mmW channel model of New York University (NYU). Hence, authors in [19] studied HBF-NOMA performance analysis based on NYU mmW channel and fixed power allocation scheme, wherein SCS-HBF-NOMA built on SIC zero forcing (ZF) scheme manifested a wider performance gap and higher computational complexity compared to FCS-HBF-NOMA. Furthermore, the extension to multistream HBF-NOMA synario was not detailed in [19]. Moreover, multistream HBF structure exploited in [20] leveraged on RF chains diversity scheme for single stream HBF-NOMA communications is simple because the signal processing did not involve combiner. Different from the existing works, this treatise proposes a simple but feasible multistream phase-zero forcing (P-ZF) SCS-HBF on the basis of unique dominant subarray matrix element extractor (DSMEE) in conjunction with a multistream FCS-HBF design solutions for mmW NOMA communications in a line of sight (LOS) of a typical urban micro (UMi) cell environment. We focus more on LOS link because in case of small cell configuration, mmW signal travels by LOS [21], which favors the exploitation of NOMA communications. Nevertheless, investigating the system's performance in a complete non-line of sight (NLOS) configuration is also significant to reveal an insight into the performance degradation arisen from the LOS link blockage, namely physical obstacles such as foliage, buildings, and mountains. The significant contributions of this treatise are recapitulated as follows:

1) We develop a multistream fully connected HBF structure articulating an explicit relationship between each RF chain at the output of digital beamformer (DB) and the corresponding antennas via analog beamformer (AB) components. Based on the developed multistream FCS-HBF structure and the existing multistream SCS-HBF structure [19], the hardware complexity arisen from the number of PSs incorporated in the multistream AB unit of HBF structure can be visualized, which was not revealed in [19]. Hence, our analyses of both the system model formulation and sum-rate maximization of the proposed scheme are carried out based on the multistream HBF structures, different from single stream system based analyses in [19].
2) In order to design an energy-efficient multistream HBF-NOMA system characterized by a SCS at both BS and users, each user's AC is optimized on the basis of singular value decomposition (SVD) [22] after which a low complexity phased and ZF aided block diagonalization schemes [13] are employed to optimize the (multistream) AB and DB of an SCS-HBF-NOMA system, respectively. Explicitly, each user's AC and their corresponding AB vectors are block diagonalized exploiting a novel DSMEE method.
3) For the purpose of benchmarking the SIC-ZF based SCS-HBF-NOMA and the optimal P-ZF FCS-HBF-NOMA in [19] to our proposed P-ZF based SCS-HBF-NOMA scheme, a simple fixed power domain NOMA is adopted.
4) Finally, computational complexities of the proposed multistream precoding and combining schemes are evaluated in terms of memory space complexity under fixed power domain NOMA assumption.

The rest of this treatise is structured as follows. Methodology to formulate the system model of the proposed downlink multiuser mmW MIMO-HBF-NOMA is elucidated in Section II. Maximization problem is developed to optimize the sum-rate in Section III. Moreover, design solutions proposed for analog combiners (ACs) as well as analog and digital precoders are thoroughly discussed. Also, the power allocation scheme is recapitulated. In Section IV, results obtained from

simulation are presented to corroborate our findings with regard to the achievable energy efficiency (EE) and sum-rate. Finally, the paper is concluded in Section V.

Notation: We exploit both the lower-case and upper-case boldface letters to represent vectors and matrices, respectively. $(\cdot)^H$, $(\cdot)^T$, $(\cdot)^{-1}$, $\mathrm{tr}(\cdot)$, $|\cdot|$, and $\|\cdot\|_F$ designate the conjugate transpose, transpose, matrix inversion, trace of a matrix, amplitude, and Frobenius norm process, respectively. $\mathbb{E}(\cdot)$ represents the expectation. Also, $\mathbf{a}(i)$, $\mathbf{A}(i)$, $\mathbf{A}(i,i)$, and $\mathbf{A}(:,j)$ represent vector a having ith row, every element, the diagonal elements, and all rows' elements in jth column of matrix A, respectively. Annotation $\mathcal{CN}(\psi, \Omega)$ represents the Gaussian distribution having mean at ψ and variance of Ω. Finally, $\mathbf{I}_{NN_s}$ represents an $NN_s \times NN_s$ identity matrix, where N and N_s denote the number of clusters and that of transmitted symbol streams, respectively.

II. METHODOLOGY

The proposed downlink of a single carrier multiuser mmW massive MIMO-NOMA system's specifications is depicted in Fig. 1, where the transmitter (BS) is equipped with transmit antennas N_t and RF chains N_t^{RF}. The BS communicates to each of the M users per cluster through the mmW channel $\mathbf{H}$. As evident in user m of Fig. 1, each user is equipped with receive antennas $N_r \geq N_s$ and N_r^{RF} $(= N_s)$ RF chains to support N_s data stream. On the other hand, the BS constitutes

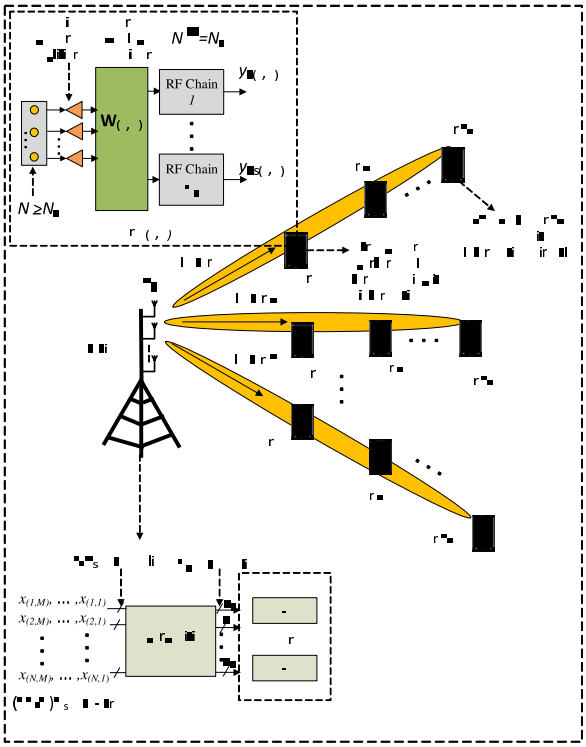

FIGURE 1. Detailed visualization of a downlink mmW multiuser MIMO-NOMA systems, where BS constitutes (FCS or SCS) HBF and N_t uniform linear array (ULA) antennas. Multiusers sharing the same unidirectional angle of arrival (AOA) are clustered together. M users per cluster can be linked by the BS simultaneously.

either FCS-HBF or SCS-HBF as manifested in Fig. 1, having $N_t^{RF} = N \cdot N_s$, where N and N_s denote the number of clusters and that of data streams, respectively. Furthermore, the AB constitutes $N_t \cdot N \cdot N_s$ and N_t PSs for FCS and SCS, respectively. As evident in Fig. 2 for FCS-HBF, each group having nth $(N \cdot N_s)$ RF chains connects to all N_t antennas via the $N \cdot N_s$ sets of PSs at the BS, whereas the corresponding SCS-HBF in Fig. 3 visualizes that each group having nth $(N \cdot N_s)$ RF chains connects to the $\mathcal{M}N_s$ sets of PSs, where $\mathcal{M}\left(= \frac{N_t}{N_t^{RF}}\right)$ denotes the sub-array antennas.

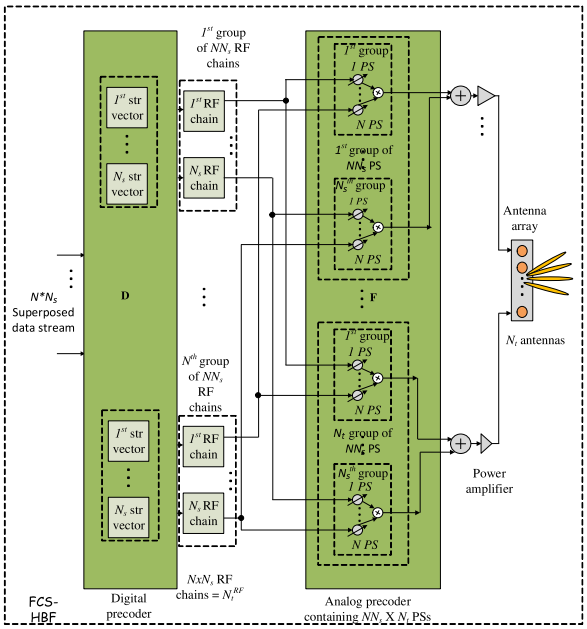

FIGURE 2. Multistream fully connected HBF structure.

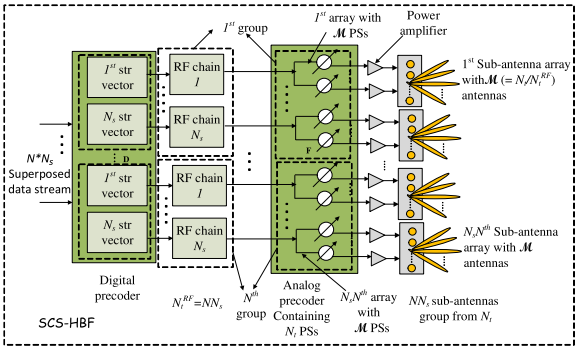

FIGURE 3. Multistream subconnected HBF structure.

Due to the limited number of RF chains in HBF system, the number of RF chains is restricted to $NN_s \leq N_t^{RF} \leq N_t$ and $N_s \leq N_r^{RF} \leq N_r$ for the BS and each user, respectively. Unlike traditional multiuser HBF system, where a user accesses BS resources through a single beam, the HBF-NOMA network can support M correlated clustered users simultaneously. Explicitly, $M \cdot N$ users can be spatially served. For the purpose of signal processing analysis,

$[NN_s \times M]$ users' data streams will be ordered for superposition coded signal $\mathbf{x}_n \in \mathbb{C}^{N_s \times 1}$ for $n = 1, \ldots, N$. Every superposition coded signal $\mathbf{x} \in \mathbb{C}^{NN_s \times 1}$ is written as $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]^T$ and then precoded exploiting $\mathbf{D} \in \mathbb{C}^{NN_s \times NN_s}$ digital precoder as depicted in Fig. 2 and 3. Then, the output of $\mathbf{D}$ is converted to RF signal exploiting analog precoder $\mathbf{F} \in \mathbb{C}^{N_t \times NN_s}$. In this regard, the RF output of the AB, namely the transmit signal $\mathbf{s}$ can be expressed as

$$\mathbf{s} = \mathbf{FDx}. \tag{1}$$

Notably, each element of $\mathbf{x}$ can be formulated as $\mathbf{x}_n = \sum_{m=1}^{M} \left(\sqrt{\frac{p_{(n,m)}}{N_s}} \cdot \mathbf{I}_{N_s}\right) \cdot \mathbf{x}_{(n,m)}$, represented as N_s superposition coded signal instigated by power domain NOMA, where $p_{(n,m)}$ means $U_{(n,m)}$ user's power coefficient allocated for multistream transmission and $\mathbf{x}_{(n,m)} \in \mathbb{C}^{N_s \times 1}$ denotes a transmitted multistream symbol for user $U_{(n,m)}$. Furthermore, the signal vector is constrained to $\mathbb{E}\left[\mathbf{x}\mathbf{x}^H\right] = \frac{P}{NN_s} \cdot \mathbf{I}_{NN_s}$, where P is the aggregate transmit power.

On the account of every AB element constituting constant amplitude PSs, the elements of $\mathbf{F}$ are normalized to $\| \mathbf{F}(i,j) \|^2 = N_t^{-1}$ for FCS and $\| \mathbf{F}(i,j) \|^2 = \mathcal{M}^{-1}$ for SCS, where $i = 1, \ldots, N_t$ and $j = 1, \ldots, NN_s$. In order to comply with the aggregate transmit power constraint, $\mathbf{D}$ is normalized to satisfy $\| \mathbf{FD} \|_F^2 = NN_s$. Under a flat fading scenario, $\mathbf{y}_{(n,m)} \in \mathbb{C}^{N_s \times 1}$ received signal at the output of RF chains of the $U_{(n,m)}$ user is written as:

$$\begin{aligned}\mathbf{y}_{(n,m)} = & \underbrace{\mathbf{W}_{(n,m)}\mathbf{H}_{(n,m)}\mathbf{F}\mathbf{D}_n\sqrt{\mathbf{P}_{(n,m)}}\mathbf{x}_{(n,m)}}_{\text{desired signal}} \\ & \underbrace{+\mathbf{W}_{(n,m)}\mathbf{H}_{(n,m)}\mathbf{F}\mathbf{D}_n \sum_{k\neq m}^{M} \sqrt{\mathbf{P}_{(n,k)}}\mathbf{x}_{(n,k)}}_{\text{intracluster interference}} \\ & \underbrace{+\mathbf{W}_{(n,m)}\mathbf{H}_{(n,m)}\mathbf{F} \sum_{l\neq n}^{N} \mathbf{D}_l \sum_{q=1}^{M} \sqrt{\mathbf{P}_{(l,q)}}\mathbf{x}_{(l,q)}}_{\text{intercluster interference}} \\ & + \underbrace{\mathbf{W}^H_{(n,m)}\mathbf{W}_{(n,m)}\mathbf{g}_{(n,m)}}_{\text{noise}},\end{aligned} \tag{2}$$

where $\mathbf{W}_{(n,m)} \in \mathbb{C}^{N_r \times N_r^{RF}}$ and $\mathbf{H}_{(n,m)} \in \mathbb{C}^{N_r \times N_t}$ represent the multistream AC and mmW MIMO channel coefficient between the user $U_{(n,m)}$ and BS, respectively. $\mathbf{D}_n \in \mathbb{C}^{NN_s \times N_s}$ indicates the corresponding N_s column vectors in $\mathbf{D}$ precoder, which boost the transmit signal towards the nth cluster. $\sqrt{\mathbf{P}_{(n,m)}} = \left(\sqrt{\frac{p_{(n,m)}}{N_s}}\right) \cdot \mathbf{I}_{N_s}$ comprises each symbol stream power and $\mathbf{g}_{(n,m)} \in \mathbb{C}^{N_r \times 1}$ represents additive white Gaussian noise (AWGN) annotated as $\sim \mathcal{CN}\left(0, \sigma^2\right)$ at N_r receive antennas of each user. It is also noted that there exists $k < m$ users and $l \neq n$ clusters.

TABLE 1. Users' channels configurations exploited in NYUSIM.

Parameters	Values
Carrier Frequency	28 $[GHz]$
RF Bandwidth	800 $[MHz]$
Scenario & Environment	UMi cell & LOS
BS to $U_{(n,m)}$ distance for cluster 1	$65.6, 83.3$ $[m]$
BS to $U_{(n,m)}$ distance for cluster 2	$56.3, 89.4$ $[m]$
Barometric pressure	1013.25 mbar
Humidity and temperature	50%, 20^oC, respectively.
Transmit power and BS height	30 $[dB]$ and 35 $[m]$, respectively.
BS and $U_{(n,m)}$ Antennas	256 and 16, respectively
Antenna Array	Universal Linear Array (ULA)
Antenna spacing	0.5λ at BS and users
AOD for Clusters 1 to 2	10^o and 20^o
AOA for Clusters 1 to 2	10^o and 20^o

NYU mmW channel model[1] for multicarrier MIMO-NOMA system in [24] has been adopted to a single carrier counterpart. Hence, single carrier channel coefficients having v multipath components between the BS and each user $U_{(m,n)}$ are formulated as

$$\begin{aligned}\mathbf{H}_{(n,m)} = & \sum_{v=1}^{V} \gamma_{(n,m,v)} e^{j\Phi_{(n,m,v)}} e^{-j2\pi f \tau_{(n,m,v)}} \\ & \times \mathbf{a}_U\left(\theta_{(n,m,v)}\right) \mathbf{a}_{BS}^H\left(\phi_{(n,m,v)}\right),\end{aligned} \tag{3}$$

where

$$\mathbf{a}_{BS}(\phi) = \frac{1}{\sqrt{N_t}}\left[1, e^{j\frac{2\pi}{\lambda}d_T sin(\phi)}, \ldots, e^{j(N_t-1)\frac{2\pi}{\lambda}d_T sin(\phi)}\right]^T \tag{4}$$

and

$$\mathbf{a}_U(\theta) = \frac{1}{\sqrt{N_r}}\left[1, e^{j\frac{2\pi}{\lambda}d_R sin(\theta)}, \ldots, e^{j(N_r-1)\frac{2\pi}{\lambda}d_R sin(\theta)}\right]^T \tag{5}$$

represent array steering vectors at the BS and user, respectively, $\gamma_{(n,m,v)}$, τ, Φ, f, d_T, d_R, $\phi_{(n,m,v)}$, and $\theta_{(n,m,v)}$ denote the amplitude of the channel gain in v^{th} path, time delay, the phase of the multipath component, carrier frequency, antenna element spacing at the BS, antenna spacing at the receiver, angle of departure (AOD), and AOA, respectively. Notably, for the purpose of achieving low antenna correlation, both d_T and d_R are set to 0.5λ, where λ denotes wavelength of the signal [25]. These parameters are extracted from the output file saved as "DirPDPinfo.mat" produced from the NYUSIM software. Table 1 depicts the mmW NOMA system configurations to generate the "DirPDPinfo.mat" output files. It is worthy to mention that achievable performance analysis for the proposed mmW HBF-NOMA system based on NYU channel model enables to guide the mobile system's designer to corroborate the realistic performance of the cutting edge system [24].

[1]For further details of NYU mmW channel model, refer to NYU 5G website [23].

III. OPTIMIZATION PROBLEM FORMULATION AND SOLUTION

The optimization problem is developed by firstly modeling the R_{sum} sum-rate of the proposed HBF-NOMA scheme as

$$R_{sum} = \sum_{n=1}^{N} \sum_{m=1}^{M} R_{(n,m)}, \tag{6}$$

where the data-rate $R_{(n,m)}$ of mth user in nth cluster is written as

$$R_{(n,m)} = \log_2 \left(\left| \mathbf{I}_{N_s} + \mathrm{SINR}_{(n,m)} \right| \right), \tag{7}$$

and

$$\mathrm{SINR}_{(n,m)} = \frac{\mathbf{P}_n \times \alpha_{(n,m)} |\tilde{\mathbf{H}}_{(n,m)} \mathbf{D}_n|^2}{I^{intra}_{(n,m)} + I^{inter}_{(n,m)} + \left| \mathbf{W}_{(n,m)} \right|^2 \sigma^2_{(n,m)}}, \tag{8}$$

where $\tilde{\mathbf{H}}_{(n,m)} \in \mathbb{C}^{N_s \times NN_s} = \mathbf{W}^H_{(n,m)} \mathbf{H}_{(n,m)} \mathbf{F}$ denotes analog channel of user $U_{(n,m)}$, $\alpha_{(n,m)}$ represents the power ration coefficient for $U_{(n,m)}$th user, intracluster interference $I^{intra}_{(n,m)}$ and intercluster interference $I^{inter}_{(n,m)}$ are formulated as $I^{intra}_{(n,m)} = \sum_{k=1}^{m-1} \mathbf{P}_n \times \alpha_{(n,k)} |\tilde{\mathbf{H}}_{(n,m)} \mathbf{D}_n|^2$ and $I^{inter}_{(n,m)} = \sum_{l \neq n}^{N} \sum_{q=1}^{M} \mathbf{P}_l \times \alpha_{(l,q)} |\tilde{\mathbf{H}}_{(n,m)} \mathbf{D}_l|^2$, respectively for $k < m$ and $n \neq l$ adjacent clusters. $\sigma^2_{(n,m)}$ denotes noise variance owing to AWGN emanating from receiver's antennas. Notably, the channel state information $\mathrm{SINR}_{(n,\Upsilon) \to (n,m)}$ written as

$$\mathrm{SINR}_{(n,\Upsilon) \to (n,m)} = \frac{\mathbf{P}_{(n,\Upsilon)} |\tilde{\mathbf{H}}_{(n,m)} \mathbf{D}_n|^2}{\sum_{k=1}^{\Upsilon-1} I^{intra}_{(n,k) \to (n,m)} + I^{inter}_{(n,m)} + \tilde{\sigma}^2_{(n,m)}} \tag{9}$$

is postulated for user $U_{(n,m)}$ to successfully decode the message of other weak users $U_{(n,\Upsilon)}$ in nth cluster, where $\Upsilon > m$, $I^{intra}_{(n,k) \to (n,m)} = \mathbf{P}_{(n,k)} |\tilde{\mathbf{H}}_{(n,m)} \mathbf{D}_n|^2$, $\mathbf{P}_{(n,\Upsilon)} = \mathbf{P}_n \times \alpha_{(n,\Upsilon)}$, and $\tilde{\sigma}^2_{(n,m)} = \left| \mathbf{W}_{(n,m)} \right|^2 \sigma^2_{(n,m)}$. A perfect SIC scheme at the strong user is assumed to eliminate the intracluster interference arisen from the feeble users and the removal of intercluster interference is guaranteed at the strongest user in each cluster via the efficient digital precoding scheme built on the basis of composite matrix constituting $\tilde{\mathbf{H}}_{(n,1)}$ effective analog channel of the strongest user in each cluster [26]. Therefore, data-rate of the strongest user can be simplified as

$$R_{(n,1)} = \log_2 \left(\left| \mathbf{I}_{N_s} + \frac{\mathbf{P}_n \times \alpha_{(n,1)} |\tilde{\mathbf{H}}_{(n,1)} \mathbf{D}_n|^2}{\tilde{\sigma}^2_{(n,m)}} \right| \right). \tag{10}$$

Maximization of the sum-rate in (6) is achievable through the effective users pairing and ordering, which is crucial for obtaining the optimal analog combiner $\check{\mathbf{W}}_{(n,m)}$, hybrid precoder $\check{\mathbf{F}} \cdot \check{\mathbf{D}}_n$ and power $\check{\mathbf{P}}_{(n,m)}$ for user $U_{(n,m)}$. Hence, we aim for maximizing the optimization problem in (11) as follows:

$$\underset{\{\mathbf{W}_n, \mathbf{F}, \mathbf{D}_n, \mathbf{P}_{(n,m)}\}}{\text{maximize}} \sum_{n=1}^{N} \sum_{m=1}^{M} R_{(n,m)}, \tag{11}$$

$$\text{subject to} |\mathbf{F}(i,j)|^2 = N_t^{-1}, \tag{12}$$

$$|\mathbf{W}_{(n,m)}(i,j)|^2 = N_r^{-1}, \tag{13}$$

$$\| \mathbf{F} [\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_N] \|_F^2 = NN_s, \tag{14}$$

$$R_{(n,m)} \geq R_{OMA}, \tag{15}$$

$$\sum_{n=1}^{N} \sum_{m=1}^{M} \mathrm{tr} \left(\mathbf{P}_{(n,m)} \right) \leq P, \tag{16}$$

$$\mathbf{P}_{(n,m)} (i,i) > 0. \tag{17}$$

Solving the objective function in (11) is non-convex owing to its subjected constraints. Therefore, we subdivide the optimization into two disjointed problems, namely the HBF and power optimization. In literature, the iterative approaches of designing optimal precoders such as alternating minimization-based algorithms [8] and orthogonal matching pursuits algorithm [4] among other iteration-based optimization schemes are capable of approaching global optimal at the expense of high computational overhead cost. Instead, non-iterative precoding leveraging on the system linearity is known to be sub-optimal to the iterative precoding [14]. Moreover, the use of non-iteration-based precoding becomes feasible and its performance enables to converge to a local optimality. For the sake of realizing a practical system having lower computational complexity in conjunction with carefully compromized performance, a best possible harmonization of the non-iteration-based precoding, combining, and a refined power allocation is highly desirable [19], [26].

In order to benchmark the existing traditional multiuser and OMA schemes, the traditional multiuser's sum-rate R^{trad}_s is expressed as [14]

$$R^{trad}_s = \sum_{n=1}^{N} \log_2 \left(\left| \mathbf{I}_{N_s} + \mathrm{SINR}^{trad}_{(n,m)} \right| \right), \tag{18}$$

where $\mathrm{SINR}^{trad}_{(n,m)} = \frac{\mathbf{P}_{(n,m)} \left| \tilde{\mathbf{H}}_{(n,m)} \mathbf{D}_n \right|^2}{\sum_{l \neq n}^{N} \mathbf{p}_{(l,m)} \left| \tilde{\mathbf{H}}_{(n,m)} \mathbf{D}_l \right|^2 + \tilde{\sigma}^2_{(n,m)}}$ and achievable data-rate of OMA scheme $R^{OMA}_{(n,m)}$ is obtained from [27], [28]

$$R^{OMA}_{(n,m)} = \frac{1}{M} \log_2 \left(\left| \mathbf{I}_{N_s} + \frac{\mathbf{P}_{(n,m)} |\tilde{\mathbf{H}}_{(n,m)} \mathbf{D}_n|^2}{\frac{1}{M} \tilde{\sigma}^2_{(n,m)}} \right| \right), \tag{19}$$

where M denotes the number of users per cluster. Based on (19), an achievable sum-rateR^{OMA}_s of OMA scheme is computed from

$$R^{OMA}_s = \sum_{n=1}^{N} \sum_{m=1}^{M} R^{OMA}_{(n,m)}. \tag{20}$$

A. USERS PAIRING AND ORDERING

The cluster head selection algorithm and user's correlation have been established to allocate users in each cluster owing to the randomly generated geometric channel model used in [18], where the AOA is randomly generated. In our scenario, the realistic channel model is generated for each pair of deterministic AOD and AOA. For effective communication, AOA is usually modelled as the AOD in LOS link. Hence, every user in the unidirectional AOA linking with the same

AOD at the BS is allocated to the same cluster. This approach is exploited in line with both the geometric channel model and realistic NYU channel model, thereby analog precoders are often designed based on array steering angle [26], [29]. For instance, clusters [1, 2] correspond to $[10^o, 20^o]$ AODs' directions, respectively are linked with $[10^o, 20^o]$ AOAs.

Therefore, M users' channels are generated and sorted in a descending order of their channel magnitudes in each cluster. The strongest user in each cluster is usually the first user of each cluster. Specific AC is designed for each user based on the user's channel coefficient and used to formulate each user's intermediate analog channel. At the BS, AB is designed from the composite intermediate analog channels of the strongest users from all clusters in order to maximize the BF gain. Then, the analog channel is computed for each user in each of the clusters. For user m in nth cluster, the analog channel is denoted as $\tilde{\mathbf{H}}_{(n,m)} = \mathbf{W}^H_{(n,m)}\mathbf{H}_{(n,m)}\mathbf{F}$.

Furthermore, analog channels are sorted based on their weights to extract the maximized HBF channel gain for each cluster, namely $\tilde{\mathbf{H}}_{(n,1)} = \max\left\{\left|\tilde{\mathbf{H}}_{(n,1)}\right|, \ldots, \left|\tilde{\mathbf{H}}_{(n,M)}\right|\right\}$ for $n = 1, \ldots, N$ clusters. Hence, the analog channel with a maximum BF gain in each cluster is selected to formulate the composite analog channel for the sake of designing the DBF employing ZF algorithm. Notably, DBF based on strongest users does not only mitigate the intercluster interference at the strongest users but can filter off a certain degree of intercluster interferences generated at the weak users owing to the fact that clustered users are paired on the basis of equal AOAs. Finally, for the purpose of aiding an SIC processing commonly deployed to mitigate intracluster interference in NOMA communications, the users' power coefficients are rationed in ascending order on the basis of their effective hybrid digital-analog channel weights denoted as $\left|\tilde{\mathbf{H}}_{(n,m)}\mathbf{D}_n\right|$.

B. PROPOSED ANALOG COMBINERS

In order to harvest large intermediate analog channel gain from each user, the analog combiner $\mathbf{W}_{(n,m)} \in \mathbb{C}^{N_r \times N_s}$ via an SVD of user $U_{(n,m)}$'s channel coefficient is proposed, which is available at the BS by limited feedback [25]. Hence, the optimized $\mathbf{W}^{full}_{(n,m)} \in \mathbb{C}^{N_r \times N_s}$ is denoted by $\check{\mathbf{W}}^{full}_{(n,m)}$ for FCS case, which is computed from the selected first N_s column vectors of the left unitary matrix $\mathbf{U}$. SVD of each user's channel is computed to obtain $\mathbf{U}$ as

$$\mathbf{U}\Sigma\mathbf{V}^H = \text{SVD}\left(\mathbf{H}_{(n,m)}\right), \tag{21}$$

where $\mathbf{V} \in \mathbb{C}^{N_t \times N_t}$ and $\Sigma \in \mathbb{R}^{N_r \times N_t}$ represent the right unitary (square) matrix and the diagonalized singular values on $\mathbf{U}$ and $\mathbf{V}$ matrices, respectively. Based on (21), FCS combiner is formulated as [22]

$$\check{\mathbf{W}}^{full}_{(n,m)} = \frac{1}{\sqrt{N_r}}\left[\mathbf{U}^{(N_r,1:N_s)}_{(n,m)}\right], \tag{22}$$

where $1 : N_s$ denotes the first N_s column vector spaces in $\mathbf{U}$. In case of the optimized SCS-AC, $\check{\mathbf{W}}^{sub}_{(n,m)} \in \mathbb{C}^{N_r \times N_s}$ can be formulated as

$$\check{\mathbf{W}}^{sub}_{(n,m)} = \mathcal{M}_r diag\left[\frac{1}{\sqrt{\mathcal{M}_r}}\mathbf{W}^{full}_{(n,m)}\right], \tag{23}$$

where $\mathcal{M}_r$ is obtained from $\mathcal{M}_r = \frac{N_r}{N_s}$.

For the sake of solving the block diagonalization constraint in (23), we postulate a selective matrix $\mathbf{G}^r_i \in \mathbb{C}^{\mathcal{M}_r \times N_r}$ that chooses $\check{\mathbf{w}}^i_{(n,m)} \in \mathbb{C}^{\mathcal{M}_r \times 1}$ dominant elements of each ith column vector of $\mathbf{W}^{full}_{(n,m)}$, where $i = 1, \ldots, N_s$. Therefore, $\mathbf{G}^r_i$ for each *ith* sub-array column is formulated as

$$\mathbf{G}^r_i = \left[\mathbf{0}_{(\mathcal{M}_r,\mathcal{M}_r(i-1))}, \mathbf{I}_{\mathcal{M}_r}, \mathbf{0}_{(\mathcal{M}_r,\mathcal{M}_r(N_s-i))}\right] \tag{24}$$

for $i = 1$ to N_s. Hence, $\check{\mathbf{w}}^i_{(n,m)} \in \mathbb{C}^{\mathcal{M} \times 1}$ is obtained from

$$\check{\mathbf{w}}^i_{(n,m)} = \frac{1}{\sqrt{\mathcal{M}_r}}\mathbf{G}^r_i\mathbf{W}^{full}_{(n,m)}\left(N_r, i\right). \tag{25}$$

Then, an optimal $\check{\mathbf{w}}^{(i,sub)}_{(n,m)} \in \mathbb{C}^{N_r \times 1}$ is ith column of N_s column vectors of $\check{\mathbf{W}}^{sub}_{(n,m)}$, which contains the non-zero dominant elements $\check{\mathbf{w}}^i_{(n,m)}$ and non-dominant elements having $(N_r - \mathcal{M}_r)$ zeros formulated as

$$\check{\mathbf{w}}^{(i,sub)}_{(n,m)} = \left[\mathbf{0}_{(1\times\mathcal{M}_r(i-1))}, \check{\mathbf{w}}^i_{(n,m)}, \mathbf{0}_{(1\times\mathcal{M}_r(N_s-i))}\right] \tag{26}$$

for $i = 1, .., N_s$ to update $\check{\mathbf{W}}^{sub}_{(n,m)}$ as:

$$\check{\mathbf{W}}^{sub}_{(n,m)} = \begin{bmatrix} \check{\mathbf{w}}^{(1,sub)}_{(n,m)} & . & . & . & 0 \\ 0 & \check{\mathbf{w}}^{(2,sub)}_{(n,m)} & . & . & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & . & . & \check{\mathbf{w}}^{(N_s,sub)}_{(n,m)} \end{bmatrix} \tag{27}$$

C. PROPOSED ANALOG PRECODERS

Analog precoder is designed on the basis of phased scheme, which is matched to $\hat{\mathbf{H}}_{(n,1)}$ intermediate analog channel of the strongest user in each cluster. The maximized weight is usually obtained from $\hat{\mathbf{H}}_{(n,1)} \in \mathbb{C}^{N_s \times N_t} = \mathbf{W}^H_{(n,1)}\mathbf{H}_{(n,1)}$ under the assumption of $\left|\hat{\mathbf{H}}_{(n,1)}\right| > \left|\hat{\mathbf{H}}_{(n,2)}\right| > \ldots > \left|\hat{\mathbf{H}}_{(n,M)}\right|$, where $\hat{\mathbf{H}}_{(n,m)} = \mathbf{W}^H_{(n,m)}\mathbf{H}_{(n,m)}$ for user $U_{(n,m)}$. Therefore, the composite of intermediate channels is expressed as

$$\widehat{\mathbf{H}}_{(n,1)} = \left[\hat{\mathbf{H}}^T_{(1,1)}, \hat{\mathbf{H}}^T_{(2,1)}, \ldots, \hat{\mathbf{H}}^T_{(N,1)}\right]^T. \tag{28}$$

In the regard of (28), analog precoder $\mathbf{F}$ is derived for the FCS precoder as

$$\mathbf{F}^{full} = \frac{1}{\sqrt{N_t}}e^{-j\angle\widehat{\mathbf{H}}^{full}_{(n,1)}}. \tag{29}$$

On the other hand, $\mathbf{F}$ for SCS is obtained via the DSMEE based phased scheme derived as follows. Firstly, we formulate intermediate AB with non-zero element $\mathbf{F}^{sub}_{int} \in \mathbb{C}^{N_t \times NN_s}$ for SCS as

$$\mathbf{F}^{sub}_{int} = \frac{1}{\sqrt{N_t}}e^{-j\angle\widehat{\mathbf{H}}^{sub}_{(n,1)}}, \tag{30}$$

where

$$\widehat{\mathbf{H}}^{sub}_{(n,1)} = \left[\left(\hat{\mathbf{H}}^{sub}_{(1,1)}\right)^T, \left(\hat{\mathbf{H}}^{sub}_{(2,1)}\right)^T, \ldots, \left(\hat{\mathbf{H}}^{sub}_{(N,1)}\right)^T\right]^T \quad (31)$$

and

$$\hat{\mathbf{H}}^{sub}_{(n,1)} = \left(\check{\mathbf{W}}^{sub}_{(n,1)}\right)^H \mathbf{H}_{(n,1)} \quad (32)$$

having non-zero elements. Explicitly,

$$\hat{\mathbf{H}}^{sub}_{(n,1)} = \max\left\{\left|\left(\check{\mathbf{W}}^{sub}_{(n,1)}\right)^H \mathbf{H}_{(n,1)}\right|, \ldots, \left|\left(\check{\mathbf{W}}^{sub}_{(n,M)}\right)^H \mathbf{H}_{(n,M)}\right|\right\} \quad (33)$$

Then, a novel DSMEE is proposed to extract $\mathcal{M}$ dominant elements from each column vector of $\mathbf{F}^{sub}_{int}$ to optimize $\mathbf{F}^{sub}$. Therefore, a selective precoder $\mathbf{G}^t_i \in \mathbb{C}^{\mathcal{M}\times N_t}$ that extracts $\check{\mathbf{f}}^{\mathcal{M}}_i \in \mathbb{C}^{\mathcal{M}\times 1}$ from $\mathbf{f}^{sub}_{(i,int)} \in \mathbb{C}^{N_t\times 1}$ is postulated, where $\mathbf{f}^{sub}_{(i,int)}$ is the ith column vector of $\mathbf{F}^{sub}_{int} \in \mathbb{C}^{N_t\times NN_s}$ to steer ith stream to their various nth cluster for N_s streams per cluster and $n = 1, \ldots, N$ clusters. Accordingly, DSMEE $\mathbf{G}^t_i$ for each ith stream of N_s streams per cluster is formulated as

$$\mathbf{G}^t_i = \left[\mathbf{0}_{(\mathcal{M},\mathcal{M}(i-1))}, \mathbf{I}_{\mathcal{M}}, \mathbf{0}_{(\mathcal{M},\mathcal{M}(NN_s-i))}\right] \quad (34)$$

for $i = 1, \ldots, NN_s$. $\check{\mathbf{f}}^{\mathcal{M}}_i$ is finally illustrated as

$$\check{\mathbf{f}}^{\mathcal{M}}_i = \frac{1}{\sqrt{\mathcal{M}}}\mathbf{G}^t_i \mathbf{f}^{sub}_{(i,int)}, \quad (35)$$

where $\frac{1}{\sqrt{\mathcal{M}}}$ is the normalized factor to satisfy constant magnitude constraint of SCS AB. Hence, $\check{\mathbf{f}}^{sub}_i = \left[\mathbf{0}_{(\mathcal{M}(i-1))}; \check{\mathbf{f}}^{\mathcal{M}}_i; \mathbf{0}_{(\mathcal{M}(NN_s-i))}\right]$ is changed for $i = 1, \ldots, NN_s$ to update the SCS ABF precoder $\check{\mathbf{F}}^{sub}_i \in \mathbb{C}^{N_t\times NN_s}$ using a column vector $\check{\mathbf{f}}^{sub}_i \in \mathbb{C}^{N_t\times 1}$, where $\check{\mathbf{f}}^{\mathcal{M}}_i \in \mathbb{C}^{\mathcal{M}\times 1}$ denotes the dominant matched filter[2] vector for ith stream of N_s streams per nth cluster. SCS AB $\check{\mathbf{F}}^{sub}_i$ is structurally represented by

$$\check{\mathbf{F}}^{sub}_i = \begin{bmatrix} \check{\mathbf{f}}^{\mathcal{M}}_1 & . & . & . & . & 0 \\ 0 & \check{\mathbf{f}}^{\mathcal{M}}_2 & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & \check{\mathbf{f}}^{\mathcal{M}}_{NN_s} \end{bmatrix}. \quad (36)$$

D. PROPOSED DIGITAL PRECODERS

In order to design the optimized DB, the designed optimized AC and that of AB are processed to implement the effective analog channel $\grave{\mathbf{H}}_{(n,m)} \in \mathbb{C}^{N_s\times NN_s}$ for each user. Hence, each user's effective analog channel is formulated as $\grave{\mathbf{H}}_{(n,m)} = \hat{\mathbf{H}}_{(n,m)}\mathbf{F}$. Afterwards, the cluster users ordering is performed to obtain the users, which possesses the highest analog channel exploiting $\grave{\mathbf{H}}_{(n,1)} = \max\left\{\left|\grave{\mathbf{H}}_{(n,1)}\right|, \ldots, \left|\grave{\mathbf{H}}_{(n,M)}\right|\right\}$ for $n = 1, \ldots, N$. Then, the composite effective analog channel $\widetilde{\mathbf{H}}_{(n,1)} \in \mathbb{C}^{NN_s\times NN_s}$ constituting various clusters' strongest analog channel $\grave{\mathbf{H}}_{(n,1)}$ is expressed as

$$\widetilde{\mathbf{H}}_{(n,1)} = \left[\grave{\mathbf{H}}^T_{(1,1)}, \grave{\mathbf{H}}^T_{(2,1)}, \ldots, \grave{\mathbf{H}}^T_{(N,1)}\right]^T. \quad (37)$$

Finally, ZF processing is implemented to obtain the multistream DB as [26]

$$\mathbf{D} = \widetilde{\mathbf{H}}_{(n,1)}\left(\widetilde{\mathbf{H}}_{(n,1)}\widetilde{\mathbf{H}}^H_{(n,1)}\right)^{-1}\boldsymbol{\beta}, \quad (38)$$

where $\boldsymbol{\beta}$ is the normalized factor to ensure total power constraint on digital baseband. $\mathbf{D} \in \mathbb{C}^{N^{RF}_t\times N\cdot N_s}$ multistream DB spatially multiplexes $N \cdot N_s$ streams through a single multistream AB denoted as $\mathbf{F} \in \mathbb{C}^{N_t\times NN_s}$.

E. POWER ALLOCATION SCHEME

We adopt both fixed power NOMA and a simple dynamic power NOMA built on the basis of large scale fading parameters to allocate users' power. Guided by the fixed power scheme in [26] and [30], total transmit power P is constrained to 1 [*watt*] and cluster users' power is configured to $p_n\left(=\frac{P}{N}\right)$. Furthermore, an aggregate summation of the clustered user power ration coefficients is restricted to unity, namely $\alpha_{(n,1)} + \alpha_{(n,2)} + \ldots + \alpha_{(n,M)} = 1$. p_n is shared among the clustered users according to $p_{(n,m)}\left(=\alpha_{(n,m)} \times p_n\right)$. Explicitly, the multistream power per nth cluster and multistream power per user $U_{(n,m)}$ are formulated as $\mathbf{P}_n\left(=\frac{p_n}{N_s}\mathbf{I}_{N_s}\right)$ and $\mathbf{p}_{(n,m)}\left(=\alpha_{(n,m)} \times \mathbf{P}_n\right)$, respectively. Therefore, power ration coefficients are constrained to satisfy $\alpha_{(n,1)} < \alpha_{(n,2)} < , \ldots, < \alpha_{(n,M)}$ owing to the magnitude order of users' hybrid (analog-digital) channel[3] in (39), represented as

$$\| \tilde{\mathbf{H}}_{(n,1)}\mathbf{D}_n \|_F > \| \tilde{\mathbf{H}}_{(n,2)}\mathbf{D}_n \|_F > \ldots > \| \tilde{\mathbf{H}}_{(n,m)}\mathbf{D}_n \|_F > \ldots > \| \tilde{\mathbf{H}}_{(n,M)}\mathbf{D}_n \|_F . \quad (39)$$

Adopting fixed power scheme to allocate transmit power for users can lead to unfair access to BS resources on the ground that users' distances to BS are randomly distributed and not fixed for each cluster. For the sake of allocating user's power on the basis of dynamic nature of the large scale fading parameters such as distance and path loss coefficient, a dynamic power allocation (DPA) scheme is introduced to leverage both inverse path loss (IPL) model and the user's distance to BS, which enables to optimize power ration coefficients. Based on the aforementioned, DPA based on distance model is formulated for the user $U_{(n,m)}$ power coefficient $\alpha^D_{(n,m)}$ as

$$\alpha^D_{(n,m)} = \frac{d^2_{(n,m)}}{\sum_{m=1}^{M} d^2_{(n,m)}} \text{ for } n = 1, \ldots, N \quad (40)$$

[2]In MIMO wireless communication system, matched filter precoder is implemented exploiting the conjugate transpose of the known user's analog channel.

[3]Hybrid (analog-digital) channel means the linear matrix product involving the effective analog channel matrix and DB matrix of a single user.

TABLE 2. Space complexity comparison.

NOMA Precoding	Multiplication	SVD	Inverse	Exponential	Addition
P-ZF FCS-HBF	$\mathcal{O}(\Omega_1)$	$\mathcal{O}(NMN_r^2N_t)$	$\mathcal{O}(N^3N_s^3)$	$\mathcal{O}(NN_sN_t)$	-
P-ZF SCS-HBF	$\mathcal{O}(\Omega_2)$	$\mathcal{O}(NN_sN_r)$	$\mathcal{O}(N^3N_s^3)$	$\mathcal{O}(NN_sN_t)$	-
SIC-ZF SCS-HBF	$\mathcal{O}(\Omega_3)$	$\mathcal{O}(\Omega_4)$	$\mathcal{O}(2N^3N_s^3)$	$\mathcal{O}(N\mathcal{M})$	$\mathcal{O}(N^3N_s^3)$

and DPA based on IPL model is also derived for the user $U_{(n,m)}$ power coefficient $\alpha_{(n,m)}^{IPL}$ as

$$\alpha_{(n,m)}^{IPL} = \frac{\sqrt{d_{(n,m)}^{c_{(n,m)}}}}{\sum_{m=1}^{M}\sqrt{d_{(n,m)}^{c_{(n,m)}}}} \quad \text{for } n = 1, \ldots, N, \tag{41}$$

where $d_{(n,m)}$ and $c_{(n,m)}$ represent the distance and pathloss coefficient for user $U_{(n,m)}$, respectively. Accordingly, the data-rate for each user is iterated deploying both (40) and (41) in turns to return the optimized power ration coefficient for the system's operation. The optimized power ration coefficient is obtained based on $\underset{\alpha_{(n,m)}^{\gamma}}{max}\left\{R_{(n,m)}\right\}$ for $\gamma \in [D, IPL]$, where D annotates distance.

F. ENERGY EFFICIENCY PERFORMANCE METRIC

Maximization of sum-rate plays a pivotal role in achieving EE, because EE of an HBF-NOMA denoted as E_{NOMA} is obtained from the ratio between the attainable sum-rate and the aggregate power consumed, namely HBF precoders' power and total transmit power. Hence, EE is formulated as [1], [18]

$$E_{NOMA} = \frac{R_{sum}}{P_{(HBF)} + P}, \tag{42}$$

G. SUM-RATE GAIN

For the purpose of clarity, the sum-rate gain (SRG) exploited in the discussion of results is measured based on the slope of sum-rate's curve. In this regard, the SRG is written as

$$SRG = \left(\frac{\Delta R_{sum}}{\Delta \text{SNR}}\right) bps/Hz/dB, \tag{43}$$

where Δ denotes difference on achievable maximum and minimum values.

H. SIC-ZF BASED SCS-HBF-NOMA JUXTAPOSING P-ZF COUNTERPART

In mmW NOMA communication downlink, SIC based optimized subconnected structure AB exploiting single stream and multistream per user transmission was derived from (30) in [19]. More explicitly, dominant element of $\mathbf{f}_n^{sub} \in \mathbf{F}^{sub}$ named $\check{\mathbf{f}}_n^{\mathcal{M}}$ was obtained from the conjugate transpose of the $\mathbf{v}_1$'s phase, mathematically $\check{\mathbf{f}}_n^{\mathcal{M}} = \frac{1}{\sqrt{\mathcal{M}}} e^{j\angle \mathbf{v}_1}$, where $\mathbf{v}_1$ is obtained from the SVD of $\bar{\mathbf{G}}_{n-1} \in \mathbb{C}^{\mathcal{M}\times\mathcal{M}} = \mathbf{G}_i^t\mathbf{G}_{n-1}$. It is worth mentioning that $\bar{\mathbf{G}}_{n-1}$ represents a submatrix of $\mathbf{G}_{n-1} \in \mathbb{C}^{N_t\times N_t}$, where $\mathbf{G}_{n-1} = \mathbf{H}_{(n,1)}^H(\mathbf{T}_{n-1})^{-1}\mathbf{H}_{(n,1)}$ for $N_s = 1$ and $\mathbf{T}_{n-1} = p_{(n,1)}\mathbf{H}_{(n,1)}\mathbf{F}\mathbf{F}^H\mathbf{H}_{(n,1)}^H$. Computational complexity cost to achieve the SIC based optimized AB was found higher than that of the phased scheme based AB deployed for the fully connected based HBF-NOMA. The higher complexity of SIC based AB arises from the high computational complexity costed to compute $\mathbf{G}_{n-1}$ matrix for $n = 1, \ldots N$. Instead, in this paper, the phased scheme postulated for the AB optimization leverages on the conjugate transpose of the strongest user's intermediate analog channel for each cluster in (30) rather than $\mathbf{v}_1$, which is easy to compute and feasible for massive MIMO LOS link. Both SIC and phased based analog precoding schemes are aided by ZF based digital precoding. Details on the computational complexity of both the two SCS-HBF-NOMA schemes and FCS-HBF-NOMA are recapitulated in the followings.

I. COMPUTATIONAL COMPLEXITY ANALYSIS

The computational complexities of the P-ZF based multistream FCS-HBF-NOMA and SCS-HBF-NOMA systems as well as the SIC-ZF based multistream SCS-HBF-NOMA [19] are evaluated in terms of memory space complexity exploiting Big-$\mathcal{O}$ notation, which also draws conjecture about the memory space consumed by the various HBF-NOMA algorithms. Complexity of mathematical operations such as multiplication operation, SVD function, inverse operation, and exponential function on matrices is considered for the memory space complexity evaluation. Updating precoder and combiner matrices as well as their derivative matrices costs trivial memory space complexities for all schemes. Hence, their updates do not contribute significantly to the complexity analysis. Memory space complexity that arises from cluster users ordering based on channel weights [26] is equal for all schemes. Therefore, the complexities induced by both matrix update and users' ordering are ignored in Table 2 illustrating upper bounds of the memory space complexity of the HBF-NOMA precoders.

1) P-ZF BASED FCS-HBF-NOMA COMPLEXITY

The proposed multistream FCS-HBF-NOMA scheme's memory space complexity is analyzed for AC, AB, and DB as follows:

- AC on the basis of SVD: AC consumes the memory space complexity of $\mathcal{O}(NMN_r^2N_t)$ for SVD computations based on $N \cdot M$ users exploiting (21).
- AB built on the phased scheme: memory space complexity of AB for FCS-HBF-NOMA costs a complexity of $\mathcal{O}(NN_sN_rN_t)$ and $\mathcal{O}(NN_sN_t)$ for multiplication and exponential function, respectively, in (29).
- DB leveraged on ZF scheme: memory space complexity of DB for FCS-HBF-NOMA consuming the complexities of $\mathcal{O}\left(\left(NMN_sN_r + N^2MN_s^2\right)N_t + 3N^3N_s^3\right)$ and

$\mathcal{O}\left(N^3N_s^3\right)$ corresponding to multiplication and inverse function arises from computing (38), respectively.

Therefore, the memory space complexity for the multistream FCS-HBF-NOMA leads to a complexity of Ω_1, $\mathcal{O}\left(NMN_r^2N_t\right)$, $\mathcal{O}\left(N^3N_s^3\right)$, and $\mathcal{O}\left(NN_sN_t\right)$ for multiplication, SVD, inverse operation, and exponential function, respectively as depicted in the second row of Table 2, where Ω_1 represents

$$\mathcal{O}\left(\left((NM+N)N_sN_r+N^2MN_s^2\right)N_t+3N^3N_s^3\right).$$

2) P-ZF BASED SCS-HBF-NOMA COMPLEXITY

The proposed multistream SCS-HBF-NOMA scheme's memory space complexity is analyzed for AC, AB, and DB as follows:

- AC on the basis of SVD and DSMEE methods: AC costs the memory space complexity of $\mathcal{O}\left(NMN_s\mathcal{M}_rN_r\right)$ and $\mathcal{O}\left(NMN_sN_r\right)$ corresponding to multiplication in (31) and (32) as well as SVD computations for $N \cdot M$ users exploiting (21), respectively.
- AB leveraged on phased scheme: memory space complexity of analog precoder for SCS-HBF-NOMA consumes a complexity of $\mathcal{O}\left(NN_sN_t\left(N_r+\mathcal{M}\right)\right)$, $\mathcal{O}\left(NN_sN_t\right)$, and $\mathcal{O}\left(N^3N_s^3\right)$ for multiplication, exponential function, and addition operations, respectively. Complexity to update the AB matrix only costs $\mathcal{O}\left(\left(\mathcal{M}+1+NN_s\right)N_t\right)$, which has been ignored in Table 2 on the ground that its matrix update does not involve mathematical operations and its degree of polynomial[4] arisen from both N_t and $\mathcal{M}$ exponents goes to one. In the same vain, a space complexity for computing fixed power ration coefficients requires $\mathcal{O}\left(N^2N_s^2\right)$ memory space complexity, which is also trivial and hence ignored in Table 2.
- DB contemplated on ZF scheme: memory space complexity of DB for SCS-HBF-NOMA spent the same complexity as in III-I2. Accordingly, the overall complexity for P-ZF based SCS-HBF-NOMA is summed for each of the mathematical operations and then presented in the third row of Table 2, where

$$\Omega_2 = \mathcal{O}\left(NMN_s\mathcal{M}_rN_r+\left(NN_sN_r+NN_s\mathcal{M}+NMN_sN_r+N^2MN_s^2\right)N_t+3N^3N_s^3\right).$$

3) SIC-ZF BASED SCS-HBF-NOMA COMPLEXITY

The memory space complexity of the SCS-HBF-NOMA scheme configured for multistream per user transmission is analyzed to benchmark the proposed subconnected structure aided HBF-NOMA scheme as follows:

- AC built on SVD and DSMEE methods: AC costs the memory space complexity as presented in III-I2.
- AB contemplated on SIC scheme: memory space complexity of AB for SCS-HBF-NOMA costs a complexity of

$$\mathcal{O}\left(\left(\left(N^2N_s+\mathcal{M}N\right)N_t^2+N^2N_sN_r+3N^2N_s^2+N^3N_s^2+\mathcal{M}^2N\right)N_t\right),$$

  $\mathcal{O}\left(\mathcal{M}^3\right)$, $\mathcal{O}\left(N^3N_s^3\right)$, and $\mathcal{O}\left(N\mathcal{M}\right)$ corresponding to multiplication, SVD, inverse operation, and exponential function, respectively.
- DB on the basis of ZF scheme: memory space complexity of DB for SIC-ZF based SCS-HBF-NOMA consumes the same complexity for ZF processing as in III-I1.

Accordingly, the overall complexity for SIC-ZF based SCS-HBF-NOMA are summed for each of the mathematical operations and then presented in the fourth row of Table 2, where

$$\Omega_3 = \mathcal{O}\left(NMN_s\mathcal{M}_rN_r+\left(N^2N_s+\mathcal{M}N\right)N_t^2+\left(N^2N_s\,N_r+3N^2N_s^2+N^3N_s^2+\mathcal{M}^2N+NMN_sN_r+N^2MN_s^2\right)N_t+3N^3N_s^3\right)$$

and $\Omega_4 = \mathcal{O}\left(NMN_sN_r+\mathcal{O}\left(\mathcal{M}^3\right)\right)$.

4) GENERAL OBSERVATION

SIC-ZF Based SCS-HBF-NOMA scheme leads to the highest complexity than the other HBF-NOMAs as the N_t and $\mathcal{M}$ becomes large owing to the second degree of polynomial in Ω_3 arisen from N_t^2 and the third degree of polynomial in Ω_4 induced from $\mathcal{M}^3$. The memory space computational complexity cost for obtaining the proposed multistream P-ZF based FCS-HBFNOMA precoders and combiners leads to lower memory space complexity compared to the proposed multistream SCS-HBF-NOMA owing to the DSMEE processing employed to block-diagonalize SCS-HBF (analog) combiner and precoder matrices. This conjecture can be well substantiated by comparing Ω_1 to Ω_2, namely Ω_2 manifests more memory space complexity than Ω_1. It is worth noting that multiplication operation articulates the highest memory space complexity than other operations. Hence, the conjecture is drawn solely from the multiplication operation.

IV. RESULTS

System parameters for LOS link are configured as in Table 1. Link level simulations of the traditional multiuser HBF and HBF-OMA configurations scenarios for M (= 2) users per cluster and N (= 2) clusters are implemented to benchmark the performance of our proposed multistream HBF-NOMA schemes using MATLAB [14]. N_t, N_t^{RF}, N_r, N_r^{RF}, N, and M are preset to 256, NN_s, 16, N_s, 2, and 2, respectively, where N_s takes the values of 1, 2, and 4 corresponding to single, two, and four-streams transmission scenarios, respectively. P_{BB}, P_{RF}, analog P_{PS}, and P_{PA} are set to 200 $[mW]$, 160 $[mW]$, 20 $[mW]$, and 40 $[mW]$, respectively [1] as well

[4]The degree of the polynomial represents the highest degree of any of the individual terms constituting the polynomial function. Explicitly, the degree of an individual term in a polynomial is the exponent of its variable, i.e. a polynomial function $NMN_t + NN_sN_rN_t^2 + N^2MN_s^2N_t^4$ having three individual terms results in four degree of polynomial because the highest exponent on N_t is four.

as a set of power coefficients $\left(\alpha_{(n,1)}, \alpha_{(n,2)}\right)$ is configured as [0.25, 0.75] as fixed power allocation (FPA) coefficients for two users per cluster in [30]. Since NYUSIM software can generate realistic users' channel coefficients, a single run simulation is conducted on the basis of system configurations for SNR values ranging between -10 to 30 $[dB]$. A simple but feasible DPA scheme proposed on the basis of both (40) and (41) results in marginal performances with that of FPA scheme. Hence, link level simulation results for FPA scheme are only reported here.

Attainable sum-rates for the proposed HBF-NOMA scenarios transmitting two and four streams per user compared with a single stream counterpart are illustrated in Fig. 4. It is observed that multistream FCS-HBF-NOMA obtains higher spatial multiplexing gain to yield higher sum-rate than that attained by the subconnected structure counterpart at SNRs ranged from 0 to 30 $[dB]$ and 10 to 30 $[dB]$ corresponding to two and four streams per user transmission case, respectively, while the single stream FCS-HBF-NOMA exhibits higher sum-rate than that gained by the SCS-HBF-NOMA counterpart at all SNRs. It is worth noting the reason for the performance trend differences between the HBF-NOMA configured for two and four streams as well as its single stream counterpart. More explicitly, this performance trend differences arise from impacts on both the number of multistream and the user's power sharing factor $\frac{\text{SNR}}{NN_s}$, resulting in a lower attainable sum-rate for multistream FCS-HBF-NOMA at low SNRs than that achieved by the subconnected counterpart. This is because, more power is required to boost fully connected AB structure for achieving full BF gain, whereas the power sharing factor arisen from $N_s = 1$ is enough to boost a reduced number of PSs in the fully connected single stream AB to attain higher BF gain than its subconnected structured counterpart at all SNRs.

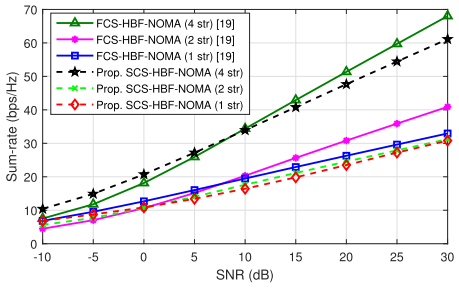

FIGURE 4. Attainable sum-rates for the proposed multistream HBF-NOMA scenarios compared with a single stream counterpart.

In order to validate the results obtained in Fig. 4, Fig. 5 illustrates attainable sum-rates of HBF-OMA and HBF traditional multiuser configured for four streams to benchmark the proposed scheme operating in LOS link. Fig. 5 portrays that the proposed SCS-HBF-NOMA scheme achieves a superior sum-rate compared to the SCS-HBF-OMA and SCS-HBF traditional multiuser configurations at all the SNRs. Notably, at the SNRs ranged between 10 and 30 $[dB]$, the HBF-OMA also achieved sum-rate gain of 2 $[bps/Hz/dB]$ using (43)

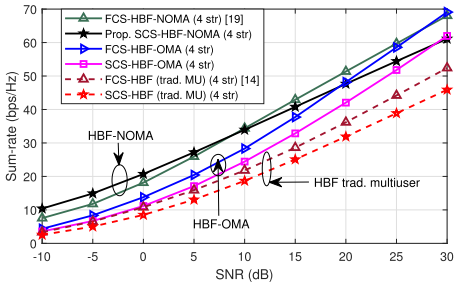

FIGURE 5. Attainable sum-rates for the HBF-OMA and HBF traditional multiuser scenarios benchmarking the HBF-NOMA schemes in LOS link.

to approach the same sum-rate as that achieved by their HBF-NOMA counterparts.

Fig. 6 shows the attainable EEs of the P-ZF based FCS-HBF-NOMA benchmarking the proposed P-ZF based SCS-HBF-NOMA scheme configured for one, two, and four streams. Results in Fig. 6 reveal that the proposed scheme yields meaningfully increased EEs as the number of data streams is increased. On the other hand, one stream FCS-HBF-NOMA offers the best possible EE compared to its two and four streams' configurations with higher sum-rates. This finding is due to the fact that the power consumed by the AB becomes higher as the number of streams is incremented. Explicitly, $N \cdot N_s \cdot N_t$ PSs are required for AB, which depends mainly on N_s configuration, whereas that required by the SCS counterpart remains N_t PSs and becomes independent of N_s. Fig. 7 illustrates EE attained by the SCS-HBF-NOMA on the basis of SIC-ZF scheme benchmarking the proposed P-ZF counterpart configured for one to four multistreams in turns. It is manifested in Fig. 7 that the proposed SCS-HBF-NOMA systems for one to four streams per user configuration yield substantively higher EE than its SIC-ZF counterpart at all SNRs. Notably, the proposed two streams assisted SCS-HBF-NOMA is capable of surpassing the EE of the single stream counterpart at SNRs ranged from 10 to 30 $[dB]$.

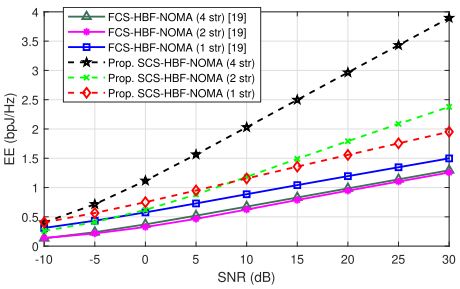

FIGURE 6. Attainable EEs for the P-ZF based (FCS and the proposed SCS).

Moreover, the sum-rate performances of the SIC-ZF based SCS-HBF-NOMA benchmarking the proposed schemes configured for two and four multistreams are portrayed in Fig. 8. Referencing the SIC-ZF based SCS-HBF-NOMA scheme at SNR of 30 $[dB]$, the proposed scheme configured for two and four streams yields performance improvements of 28 % and 33 %, respectively. At low to medium SNRs ranged from -10 to 10 $[dB]$, the proposed scheme manifests a sub-

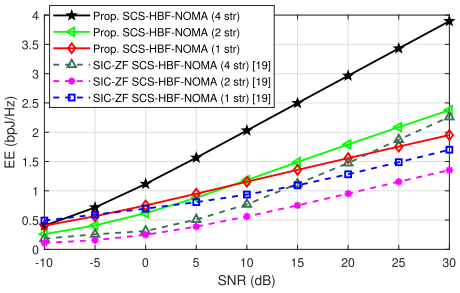

FIGURE 7. Achievable EEs for the proposed SCS-HBF-NOMA scenarios having two and four streams compared with a SIC-ZF counterpart.

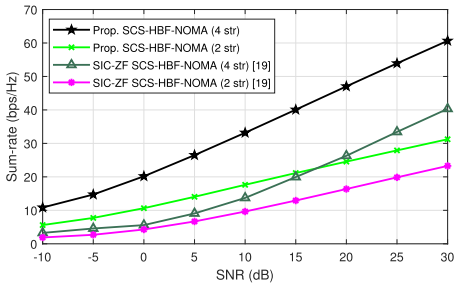

FIGURE 8. Achievable sum-rates for the proposed SCS-HBF-NOMA scenarios having two and four streams compared with a SIC-ZF counterpart.

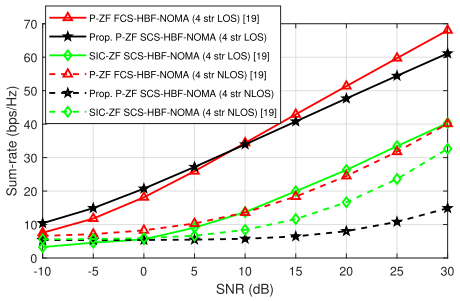

FIGURE 9. Attainable sum-rate of the various HBF-NOMA schemes configured for four streams per user in a complete LOS and NLOS links.

stantively higher multiplexing gain compared to the SIC-ZF counterpart. Furthermore, at SNR of 30 [*dB*], a multiplexing gain of 46 % is obtained by the proposed scheme, which is slightly higher than that of 43 % yielded by the SIC-ZF counterpart. Therefore, an explicit conjecture can be drawn from the performance trend that the proposed scheme is considered as a better candidate operating at low to high SNRs.

In order to reveal an insight into the performance of the P-ZF based multistream HBF-NOMA operated in NLOS link, ''DirPDPinfo.mat'' is generated for NLOS link for two users in clusters one and two, which are located at [27, 78 *m*] and [20, 133 *m*] distances to BS in conjunction with their corresponding azimuth AOAs of 50^o and 60^o, respectively [19], [31]. The NLOS link-level simulations are implemented for the P-ZF based FCS-HBF-NOMA and SCS-HBF-NOMA schemes as well as SCS-HBF-NOMA built on SIC-ZF algorithm [19] to benchmark their performances in LOS environment as illustrated in Fig. 9. It is worth noting that the four streams aided SCS-HBF-NOMA built on P-ZF processing results in the worst performance in NLOS link. Specifically, referencing sum-rates in the LOS link, FCS-HBF-NOMA and SCS-HBF-NOMA built on P-ZF schemes as well as SCS-HBF-NOMA built on SIC-ZF scheme lead to performances degradation of [59, 82, 42] % at SNR of 10 [*dB*] and [42, 75, 18] % at SNR of 30 [*dB*] in NLOS link, respectively. It is inferred from the results that a good LOS link is inevitable to deploy P-ZF scheme for SCS-HBF-NOMA system exploiting multistreams per user transmission.

V. CONCLUSION

In this paper, the sum-rate maximization problems in a mmW NOMA system operating in LOS environment have been investigated. A hybrid precoding and analog combining schemes capable of supporting multistream per user have been designed for both FCS and SCS configurations under RF chains and total transmit power constraints. Based on it, a FPA scheme have been exploited for the study and a simple DPA technique was also introduced. Substantially reduced computational complexity of the proposed scheme was explicitly corroborated via the memory space complexity analyses. Link-level simulation results manifested that HBF-NOMA attains its sum-rate higher than HBF-OMA and traditional multiuser systems. Moreover, the proposed multistream SCS-HBF-NOMA leveraging P-ZF precoding scheme yielded higher EE than the FCS-HBF-NOMA. Our findings also articulated the significance of dominant LOS link for an effective mmW HBF-MIMO-NOMA communications induced from highly correlated cluster users. A meaningful insight into SNRs requirements yielding higher attainable sum-rates for the proposed multistream SCS-HBF-NOMA compared to the fully connected counterpart has been manifested. Lastly, the proposed SCS-HBF-NOMA scheme is strongly recommended as an energy efficient BF-NOMA scheme for post-5G wireless communication.

REFERENCES

[1] W. Hao, M. Zeng, Z. Chu, and S. Yang, ''Energy-efficient power allocation in millimeter wave massive MIMO with non-orthogonal multiple access,'' *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 782–785, Dec. 2017.
[2] W. Hao, G. Sun, F. Zhou, D. Mi, J. Shi, P. Xiao, and V. C. M. Leung, ''Energy-efficient hybrid precoding design for integrated multicast-unicast millimeter wave communications with SWIPT,'' *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 10956–10968, Nov. 2019.
[3] S. He, C. Qi, Y. Wu, and Y. Huang, ''Energy-efficient transceiver design for hybrid sub-array architecture MIMO systems,'' *IEEE Access*, vol. 4, pp. 9895–9905, 2016.
[4] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Jr., ''Spatially sparse precoding in millimeter wave MIMO systems,'' *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2013.
[5] R. Rajashekar and L. Hanzo, ''Hybrid beamforming in mm-wave MIMO systems having a finite input alphabet,'' *IEEE Trans. Commun.*, vol. 64, no. 8, pp. 3337–3349, Aug. 2016.
[6] J.-C. Chen, ''Hybrid beamforming with discrete phase shifters for millimeter-wave massive MIMO systems,'' *IEEE Trans. Veh. Technol.*, vol. 66, no. 8, pp. 7604–7608, Aug. 2017.
[7] X. Gao, L. Dai, S. Han, I. Chih-Lin, and R. W. Heath, Jr., ''Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays,'' *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, Apr. 2016.
[8] X. Yu, J.-C. Shen, J. Zhang, and K. B. Letaief, ''Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems,'' *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 485–500, Apr. 2016.

[9] A. Alkhateeb and R. W. Heath, Jr., "Gram Schmidt based greedy hybrid precoding for frequency selective millimeter wave MIMO systems," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 3396–3400.

[10] X. Yu, J. Zhang, and K. B. Letaief, "Alternating minimization for hybrid precoding in multiuser OFDM mmWave systems," in *Proc. 50th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2016, pp. 281–285.

[11] Z. Wang, M. Li, Q. Liu, and A. Lee Swindlehurst, "Hybrid precoder and combiner design with low-resolution phase shifters in mmWave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 256–269, May 2018.

[12] A. Li, C. Masouros, and M. Sellathurai, "Analog–digital beamforming in the MU-MISO downlink by use of tunable antenna loads," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3114–3129, Apr. 2018.

[13] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 653–656, Dec. 2014.

[14] W. Ni and X. Dong, "Hybrid block diagonalization for massive multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 201–211, Jan. 2016.

[15] S. M. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 721–742, 2nd Quart., 2017.

[16] M. Liaqat, K. A. Noordin, T. A. Latef, and K. Dimyati, "Power-domain non orthogonal multiple access (PD-NOMA) in cooperative networks: An overview," *Wireless Netw.*, vol. 26, no. 1, pp. 181–203, Jan. 2020.

[17] W. K. New, C. Y. Leow, K. Navaie, and Z. Ding, "Robust non-orthogonal multiple access for aerial and ground users," *IEEE Trans. Wireless Commun.*, vol. 19, no. 7, pp. 4793–4805, Jul. 2020.

[18] L. Dai, B. Wang, M. Peng, and S. Chen, "Hybrid precoding-based millimeter-wave massive MIMO-NOMA with simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 1, pp. 131–141, Jan. 2019.

[19] A. A. Badrudeen, C. Y. Leow, and S. Won, "Performance analysis of hybrid beamforming precoders for multiuser millimeter wave NOMA systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 8739–8752, Aug. 2020.

[20] A. A. Badrudeen, C. Y. Leow, and S. Won, "Sub-connected structure hybrid precoding for millimeter-wave NOMA communications," *IEEE Wireless Commun. Lett.*, vol. 10, no. 6, pp. 1334–1338, Jun. 2021.

[21] M. Rouse, T. Slattery, and K. Finnell. *Millimeter Wave (MM Wave)*. Accessed: Jan. 10, 2022. [Online]. Available: https://searchnetworking.techtarget.com/definition/millimeter-wave-MM-wave

[22] A. Li and C. Masouros, "Hybrid precoding and combining design for millimeter-wave multi-user MIMO based on SVD," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.

[23] (Jan. 14, 2021). *NYUSIM: The Open Source 5G and 6G Channel Model Simulator Software*. Accessed: Jan. 10, 2022. [Online]. Available: https://wireless.engineering.nyu.edu/nyusim-5g-and-6g

[24] S. Sun, G. R. MacCartney, and T. S. Rappaport, "A novel millimeter-wave channel simulator and applications for 5G wireless communications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–7.

[25] A. Alkhateeb, G. Leus, and R. W. Heath, Jr., "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.

[26] M. A. Almasi, M. Vaezi, and H. Mehrpouyan, "Impact of beam misalignment on hybrid beamforming NOMA for mmWave communications," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4505–4518, Jun. 2019.

[27] B. Kimy, S. Lim, H. Kim, S. Suh, J. Kwun, S. Choi, C. Lee, S. Lee, and D. Hong, "Non-orthogonal multiple access in a downlink multiuser beamforming system," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, Nov. 2013, pp. 1278–1283.

[28] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6010–6023, Aug. 2016.

[29] Z. Ding, P. Fan, and H. V. Poor, "Random beamforming in millimeter-wave NOMA networks," *IEEE Access*, vol. 5, pp. 7667–7681, 2017.

[30] M. A. Almasi and H. Mehrpouyan, "Non-orthogonal multiple access based on hybrid beamforming for mmWave systems," in *Proc. IEEE 88th Veh. Technol. Conf. (VTC-Fall)*, Aug. 2018, pp. 1–7.

[31] K. Haneda *et al.*, "5G 3GPP-like channel models for outdoor urban microcellular and macrocellular environments," in *Proc. IEEE 83rd Veh. Technol. Conf. (VTC Spring)*, May 2016, pp. 1–7.

ABDULAHI ABIODUN BADRUDEEN (Member, IEEE) received the B.Eng. degree from the Federal University of Technology Minna, in 2006, the M.Tech. degree from the Ladoke Akintola University of Technology, Ogbomoso, in 2014, and the Ph.D. degree from Universiti Teknologi Malaysia, in 2021. He was registered as an Engineer by the Council for the Regulation of Engineering in Nigeria (COREN) in 2016. He is currently an Academic Staff with the Department of Computer Engineering, Federal Polytechnic Ede. His research interests include MIMO, mmWave communications, NOMA, microcontroller-based control systems, and machine learning aided post-5G.

CHEE YEN LEOW (Senior Member, IEEE) received the B.Eng. degree in computer engineering from Universiti Teknologi Malaysia (UTM), in June 2007, and the Ph.D. degree in wireless communications from Imperial College London, in September 2011.

He is currently an Associate Professor with the School of Electrical Engineering, Faculty of Engineering; and a Research Fellow with the Wireless Communication Centre, UTM. He is also the Secretary of the IMT and Future Networks Working Group under the Malaysian Technical Standards Forum Berhad to accelerate the adoption of 5G IMT-2020 in Malaysia. He also regularly conducts short courses on 4G and 5G for the telecommunication industry. His current research interests include non-orthogonal multiple access, drone communication, intelligent surfaces, advanced MIMO, millimeter wave communication, and prototype development using software-defined radio, for beyond 5G and the Internet of Things applications.

Dr. Leow is a registered Chartered Engineer (C.Eng.) of the Engineering Council U.K. His IEEE journal articles won the IEEE Malaysia Comsoc/VTS Joint Chapter's Best Paper Award in 2016, 2017, and 2021; and the IEEE Malaysia AP/MTT/EMC Joint Chapter's Best Paper Award in 2017, 2018, and 2020. He is among the pioneers for 5G initiatives in Malaysia to promote 5G research and development collaboration between industry and academia.

SEUNGHWAN WON (Senior Member, IEEE) received the B.S. and M.S. degrees in radio science and engineering from Korea University, Seoul, Republic of Korea, in 1999 and 2001, respectively, and the Ph.D. degree from the Communications Research Group, School of Electronics and Computer Science, University of Southampton, U.K. He was a Research Engineer with the Mobile Communication Technology Research Laboratory, LG Electronics Research and Development, from January 2001 to September 2004. From 2004 to 2008, he conducted research at the Communications Research Group, School of Electronics and Computer Science, University of Southampton. Upon completing his Ph.D., he returned to his native South Korea and joined Samsung. Then, he was involved in designing commercial mobile station modem. In 2013, he was appointed as an Associate Professor. He is currently teaching and conducting research with the University of Southampton Malaysia, Johor, Malaysia. His major research interests include diverse deep learning aided mobile communications associated with synchronization, multiuser MIMO and hybrid beamforming in mmW mobile communications, the IoT and UAV mobile communications, and intelligent multi-copter design. He published a lot of articles in these research fields and secured 20 U.S. and lots of Korean patents. He was a recipient of the 2004 State Scholarship of the Information and Telecommunication National Scholarship Program, Ministry of Information and Communication (MIC), Republic of Korea.