This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2022.3143221, IEEE Journal on Selected Areas in Communications

HDMA: Holographic-Pattern Division Multiple Access

Ruoqi Deng, *Student Member, IEEE*, Boya Di, *Member, IEEE*, Hongliang Zhang, *Member, IEEE*, and Lingyang Song, *Fellow, IEEE*

***Abstract*—The next generation wireless communications aiming at enhancing capacity and massive connectivity significantly over high-frequency bands urge the development of novel multiple access technologies. In this paper, we propose a new type of space-division multiple access (SDMA), called holographic-pattern division multiple access (HDMA). We develop the principle for HDMA with the main idea of mapping the intended signals for receivers to a superposed holographic pattern. The multi-user holographic beamforming scheme for HDMA is then presented. Based on the theoretical analysis, we find that there exists an optimal holographic pattern such that the sum rate with simple zero-forcing precoding can achieve the asymptotic capacity of the HDMA system. Simulation results verify the theoretical analysis and show that the HDMA scheme outperforms the traditional SDMA scheme in terms of both the cost-efficiency and the sum rate.**

***Index Terms*—Holographic-pattern division multiple access, Reconfigurable holographic surface, Asymptotic capacity analysis, Holographic pattern design**

I. Introduction

The space-division multiple access (SDMA) scheme is envisioned as one of the promising candidates to handle the exponentially increasing data transmission demands by exploiting the spatial diversity [1]. It was first proposed in multi-BS systems by exploiting spatial diversity among different BSs. Based on different spatial signatures of the BSs, different BSs can transmit signals in the same time slots without interfering with each other [2]. Evolving from multi-BS systems, multiple-input multiple-output (MIMO) systems based on the SDMA scheme have been proposed [3], where a BS simultaneously transmits multiple beams in different directions [4]. By scaling up MIMO by orders of magnitude, massive MIMO systems with large-scale phased arrays have drawn great attention due to their capabilities of highly directional beamforming and significant capacity enhancement [5], [6].

Nowadays, due to the enormous growth in the number of mobile devices [7], the next generation wireless communications look forward to enhancing capacity and massive connectivity significantly over high-frequency bands through ultra-massive MIMO systems [8]. However, the inherent defects of phased arrays limit the future development of ultra-massive MIMO systems. Specifically, considering the tradeoff between manufacturing difficulty and radiation performance, the element spacing of a phased array is usually no less than half wavelength [9], which restricts its directive gain for a given physical dimension. In addition, as the physical dimensions of phased arrays relying on costly components (such as phase shifters) scale up, the implementation of ultra-massive MIMO systems in practical engineering becomes prohibitive from both cost and power consumption perspectives [10]. Hence, there is an urgent need for developing novel multiple access technologies to meet the exponentially increasing data demand while covering the shortages of the traditional SDMA scheme.

Owing to the recent breakthrough of the reconfigurable metamaterial-based antennas, it is now possible to regulate electromagnetic waves via software instead of costly hardware components [11], [12]. Different from phased arrays, metamaterial-based antennas have compact structures where the element spacing is no larger than one quarter wavelength [13], enabling a higher directive gain in a smaller physical dimension. As one of the representative metamaterial antennas, reconfigurable holographic surfaces (RHSs) inlaid with numerous metamaterial radiation elements serving antennas integrated with transceivers have been proposed [14]. Specifically, the feeds of the RHS are embedded in the bottom layer of the RHS to generate the incident electromagnetic waves, enabling an ultra-thin structure. The RHS utilizes the metamaterial radiation elements to construct a *holographic pattern* based on the holographic interference principle [15]. Each element can thus control the *radiation amplitude* of the incident electromagnetic waves electrically to generate desired directional beams [16].

It is worth mentioning that RHSs are different from another hardware technology for wireless communication enhancement called reconfigurable intelligent surfaces (RISs) [17], which can reflect the incident signals and generate directional beams towards receivers directly. Unlike the feeds of RHSs embedded in meta-surfaces [18], the feeds of RISs are on the outside of the meta-surface due to their reflection characteristic [19]. Due to their different physical structures, RISs are widely used as a relay such that two channels (i.e., the channel from the transmitter to the RIS and the channel from the RIS to the receiver) should be considered in RIS-aided communications [20]. Differently, RHSs can serve as transmit and receive antennas directly, and only one channel (i.e., the channel between the RHS and the receiver/transmitter) needs to be considered in RHS-aided communications. Benefitting from the above characteristics, RHSs provide a powerful yet lightweight solution to fulfill the challenging visions in next

Manuscript received August 14, 2021; revised November 6, 2021; accepted December 17, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61829101, 61941101, and 61911540076, and in part by Beijing Natural Science Foundation under Grant L212027. *(Corresponding author: Lingyang Song)*

Ruoqi Deng, Boya Di, Hongliang Zhang, and Lingyang Song are with State Key Laboratory of Advanced Optical Communication Systems and Networks, Department of Electronics, Peking University, Beijing, China (email: ruoqi.deng@pku.edu.cn; boya.di@pku.edu.cn; hongliang.zhang92@gmail.com; lingyang.song@pku.edu.cn).

Hongliang Zhang is also with Department of Electrical Engineering, Princeton University, NJ, USA.

generation wireless communications and have attracted wide attention from industry [21]. Two commercial prototypes of RHSs have already been proposed by Pivotal Commwave and Kymeta. Pivotal Commwave has provided customized RHS systems from 1GHz to 70GHz to extend the coverage of cellular communications [22]. Kymeta together with Microsoft has developed the commercial RHS for satellite communications to respond to natural disasters [23].

In this paper, we propose a new type of SDMA, called holographic-pattern division multiple access (HDMA) by utilizing the superposition of different holographic patterns to serve multiple users. Specifically, in the HDMA, the transmitter superposes all holographic patterns corresponding to the receivers in different directions to generate a superposed holographic pattern, and thus, the transmitted data is mapped onto the superposed holographic pattern. When the electromagnetic waves generated by the feeds excite this holographic pattern, the signals intended for the receivers can be differentiated, i.e., the RHS can generate multiple desired directional beams towards different receivers. The directive gain brought by HDMA can be further improved by the compact element spacing of the RHS, implying that the HDMA system has more potential in enhancing capacity and massive connectivity compared with the traditional SDMA system.

Based on the principle of the HDMA, we analyze the asymptotic capacity and the sum rate of an HDMA multi-user wireless communication system. An optimal holographic pattern design scheme is thus required for asymptotic capacity achievement. This is a non-trivial task due to the following two reasons. For one thing, the holographic pattern design is based on an amplitude-controlled method, which is different from the traditional SDMA system realized by a phase-controlled method. For the other, the coupling between all radiation elements with the propagating reference wave complicates the theoretical analysis of the HDMA system as well as the holographic pattern design.

By addressing the above challenges, we contribute to the research on the HDMA wireless communication system in the following ways.

- We present the concept of HDMA. Specifically, the principle and multi-user holographic beamforming scheme for HDMA are illustrated. An HDMA wireless communication system is then considered where a BS equipped with an extremely large-scale RHS transmits signals to multiple users via the HDMA scheme. We analyze the asymptotic capacity and the sum rate with ZF precoding of the HDMA wireless communication system.
- We obtain two important results of the HDMA scheme. *First*, a closed-form optimal holographic pattern to achieve the asymptotic capacity of the HDMA wireless communication system is derived (Proposition 2 and Remark 2). *Second*, the derived optimal holographic pattern can make the sum rate with simple ZF precoding achieve the asymptotic capacity (Proposition 6). In addition, the closed-form lower and upper bounds of the data rate in a special single-user communication case are also derived.
- Simulation results verify the theoretical analysis about the capacity and the sum rate of the HDMA wireless communication system. It also shows that the HDMA wireless communication system outperforms the traditional SDMA system in terms of capacity and the maximum number of accessed users.

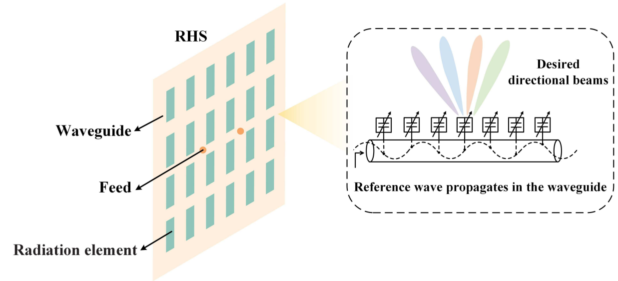

Fig. 1. Physical structure of RHS

The rest of this paper is organized as follows. In Section II, we introduce the basics of an RHS and the principle of HDMA. In Section III, we introduce an HDMA wireless communication system and give the channel model. In Section IV, we present the multi-user holographic beamforming scheme and the asymptotic capacity of the HDMA system. In Section V, we derive a closed-form optimal holographic pattern to achieve the asymptotic capacity. In Section VI, based on the derived optimal holographic pattern, we evaluate the performance of the HDMA system. The simulation results are presented in Section VII. Finally, the conclusions are drawn in Section VIII.

Notation: Scalars are denoted by italic letters, vectors and matrices are denoted by bold-face lower-case and upper-case letters, respectively. For a complex scalar x, x^* denotes its conjugate, $|x|$ denotes its modulus, and $\mathrm{Re}(x)$ denotes its real part. For a vector $\boldsymbol{v}$, $\boldsymbol{v}^T$ denotes its transpose, $|\boldsymbol{v}|$ denotes its Euclidean norm, and $\mathrm{diag}\{\boldsymbol{v}\}$ denotes the diagonal matrix whose diagonal element is the corresponding element in $\boldsymbol{v}$. For a matrix $\mathbf{G}$, $\mathbf{G}^H$ denotes its conjugate transpose, and $[\mathbf{G}]_{i,j}$ denotes the element in the i-th row and the j-th column. $\mathbf{I}_L$ denotes the identity matrix of size L. Furthermore, $\mathbb{E}(\cdot)$ and $\mathrm{Tr}(\cdot)$ denote the expectation and the trace operator, respectively.

II. Holographic-Pattern Division Multiple Access

In this section, we introduce the structure and properties of an RHS, based on which the principle of HDMA is illustrated.

A. Basics of an RHS

An RHS is a leaky-wave antenna mainly consisting of three parts, i.e., K feeds, a waveguide, and N sub-wavelength metamaterial radiation elements. Specifically, as shown in Fig. 1, the feeds are embedded in the bottom layer of the RHS and generate the incident electromagnetic waves, which are also called reference waves, carrying intended signals for users [18]. The waveguide playing the role of guiding structure then guides the reference wave to propagate on it. During the propagation process, the reference wave radiates its energy through the radiation elements into free space and the radiated wave is also known as the leaky wave [21], where

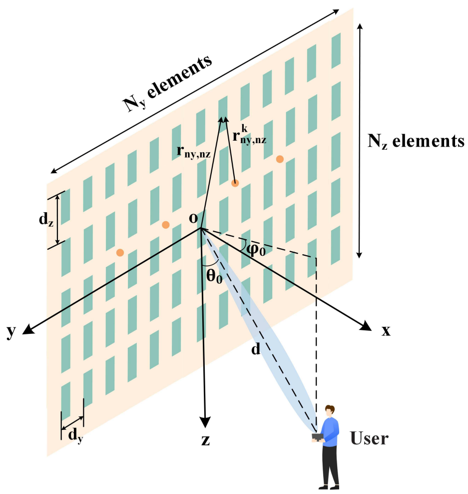

Fig. 2. Geometrical relation between the extremely large-scale RHS and object beam

the waveform of the leaky wave is the same as that of the transmit signal.

Since the electromagnetic response of the radiation elements can be intelligently controlled via a simple diode-based amplitude controller, the RHS utilizes the radiation elements to construct a holographic pattern, which records the interference between the reference wave and the desired object wave [15]. The leaked wave from each element can then be shaped independently according to the holographic pattern, and the superposition of the leaked waves from different elements finally generates the desired directional beams.

Therefore, the basic working principles and specific implementation of an RHS are different from those of a phased array. As for basic working principles, the traditional phased array realizes beamforming by controlling the phase of the electromagnetic wave propagating along each antenna of the array. Differently, the RHS can achieve amplitude-controlled beamforming by controlling the radiation amplitude of the reference wave propagating along the waveguide. As for specific implementations, unlike the phased array with complex phase-shifting circuits with numerous phase shifters, the metamaterial radiation elements of the RHS can be fabricated by loading RF switches or tunable materials rather than costly hardware components. In addition, the phased array utilizes the method of parallel feeding where each radiation element has an equal long feeding line, thereby providing in-phase feed. The RHS utilizes the method of series feeding where radiation elements are located progressively farther and farther away from the feed point. The reference wave generated by the feeds excites the radiation elements one by one.

B. Principle of HDMA

The main characteristic of the HDMA is to map the transmitted data onto a holographic pattern according to the holographic interference principle. Utilizing the holographic pattern, the RHS can control the radiation amplitude of the leaky wave at each element effectively to obtain the desired signal transmission directions. The specific principles of constructing the holographic pattern and realizing HDMA are elaborated as below.

1) Holographic Pattern Construction: As shown in Fig. 2, we adopt the Cartesian coordinate where the $y-z$ plane coincides with the RHS, and the x-axis is vertical to the RHS. The RHS is centered at the origin. The number of radiation elements along the y-axis and z-axis is denoted as N_y and N_z, respectively, satisfying $N = N_y \times N_z$. The inter-element spacing along the y-axis and z-axis is denoted as d_y and d_z, respectively. For notational convenience, it is assumed that both N_y and N_z are odd numbers. Therefore, the position vector of the (n_y, n_z)-th element is $\mathbf{r}_{n_y,n_z} = [0, n_y d_y, n_z d_z]^T$, where $n_y = \{0, \pm1, \pm2, \cdots, \pm\frac{N_y-1}{2}\}$ and $n_z = \{0, \pm1, \pm2, \cdots, \pm\frac{N_z-1}{2}\}$.

At the (n_y, n_z)-th radiation element, the reference wave generated by feed k and the wave propagating in free space with a direction of (θ_0, φ_0) can be given by [24]

$$\Psi_{ref}(\mathbf{r}^k_{n_y,n_z}) = \exp(-j\mathbf{k}_s \cdot \mathbf{r}^k_{n_y,n_z}), \tag{1}$$

$$\Psi_{obj}(\mathbf{r}_{n_y,n_z}, \theta_0, \varphi_0) = \exp(-j\mathbf{k}_f \cdot \mathbf{r}_{n_y,n_z}), \tag{2}$$

where $\mathbf{k}_s$ is the propagation vector of the reference wave, $\mathbf{r}^k_{n_y,n_z}$ is the distance vector from the feed k to the (n_y, n_z)-th radiation element, and $\mathbf{k}_f$ is the desired directional propagation vector in free space. The interference between the reference wave and the object wave is defined as

$$\Psi_{intf}(\mathbf{r}^k_{n_y,n_z}, \theta_0, \varphi_0) = \Psi_{obj}(\mathbf{r}_{n_y,n_z}, \theta_0, \varphi_0)\Psi^*_{ref}(\mathbf{r}^k_{n_y,n_z}). \tag{3}$$

The information contained in Ψ_{intf}, which is also called holographic pattern, is supposed to be recorded by the radiation elements. When the holographic pattern is excited by the reference wave, we have $\Psi_{intf}\Psi_{ref} \propto \Psi_{obj}|\Psi_{ref}|^2$, and thus, the wave propagating in the direction (θ_0, φ_0) is generated.

To construct the holographic pattern given in (3), each radiation element controls the radiation amplitude of the reference wave instead of conventional phase shifting. Specifically, the phase of the reference wave at each element is determined by the position of the element as given in (1). The radiation elements whose reference waves are in phase with the object wave are tuned to radiate strongly (large amplitude), while the radiation elements that are out of phase are detuned so as not to radiate (small amplitude) [24].

Note that the real part of the interference (i.e., Re$[\Psi_{intf}]$), i.e., cosine value of the phase difference between the reference wave and the object wave, decreases as the phase difference grows, which exactly meets the above amplitude control requirements. Therefore, Re$[\Psi_{intf}]$ can represent the radiation amplitude of each element. To avoid negative value, Re$[\Psi_{intf}]$ is normalized to $[0, 1]$. The holographic pattern $\boldsymbol{m}$ to generate the beam propagating in the direction (θ_0, φ_0) can then be parameterized mathematically by

$$\boldsymbol{m}(\theta_0, \varphi_0) = \frac{\mathrm{Re}[\Psi_{intf}(\theta_0, \varphi_0)] + 1}{2}, \tag{4}$$

and thus, the radiation amplitude of each element in the holographic pattern $\boldsymbol{m}(\theta_0, \varphi_0)$ is

$$m(\mathbf{r}^k_{n_y,n_z}, \theta_0, \varphi_0) = \frac{\mathrm{Re}[\Psi_{intf}(\mathbf{r}^k_{n_y,n_z}, \theta_0, \varphi_0)] + 1}{2}. \tag{5}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2022.3143221, IEEE Journal on Selected Areas in Communications

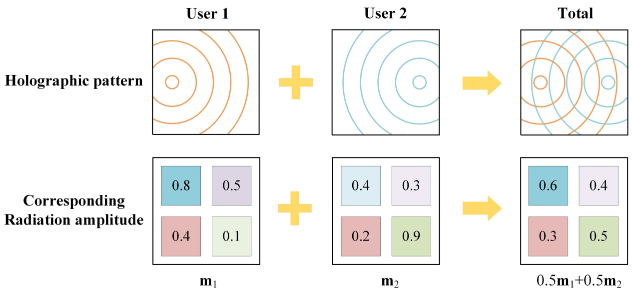

Fig. 3. Illustration of HDMA

2) Principle of HDMA: The HDMA utilizes the superposition of the holographic pattern to map all transmitted data to a single holographic pattern on the RHS, and thus, the signals intended for the users can be differentiated, i.e., the RHS can generate multiple desired directional beams pointing to the users. Specifically, the holographic pattern of the RHS can be calculated as a weighted summation of the radiation amplitude distribution corresponding to each object beam according to (5).

Denote the users served by the RHS as $\mathcal{L} = \{1, 2, \cdots, L\}$, and their zenith and azimuth angles in terms of the origin as θ_l and φ_l, respectively. The holographic pattern $\boldsymbol{m}$, i.e., the normalized radiation amplitude of each radiation element can then be given by

$$m_{n_y,n_z} = \sum_{l=1}^{L}\sum_{k=1}^{K} a_{l,k} m(\mathbf{r}^k_{n_y,n_z}, \theta_l, \varphi_l), \tag{6}$$

where $a_{l,k}$ is the amplitude ratio for the beam pointing to user l from feed k satisfying $\sum_{l=1}^{L}\sum_{k=1}^{K} a_{l,k} = 1$. This constraint is set to guarantee that the radiation amplitude of each RHS element m_{n_y,n_z} lies in [0,1].

An illustrative example for HDMA: Fig. 3 shows an example of data mapping according to the HDMA. For simplicity, we assume that the BS transmits signals to two users via an RHS[1]. The feeds of the RHS are assumed to be located close to the origin, where $\mathbf{r}^k_{n_y,n_z} \approx \mathbf{r}_{n_y,n_z}, \forall k$. Denote the zenith and azimuth angles of the user l $(l = 1, 2)$ as (θ_l, φ_l). Based on (5), we have $m(\mathbf{r}^1_{n_y,n_z}, \theta_l, \varphi_l) = m(\mathbf{r}^2_{n_y,n_z}, \theta_l, \varphi_l) = \cdots = m(\mathbf{r}^K_{n_y,n_z}, \theta_l, \varphi_l) = m(\mathbf{r}_{n_y,n_z}, \theta_l, \varphi_l)$, and thus, the index k can be omitted. The normalized radiation amplitude of each radiation element in the holographic pattern $\boldsymbol{m}_l$ corresponding to user l can then be denoted as $\{m(\mathbf{r}_{n_y,n_z}, \theta_l, \varphi_l)\}$. In the HDMA, user 1 and user 2 are multiplexed on a holographic pattern, i.e., the RHS maps the intended signals for these two users onto a single holographic pattern superposed by holographic patterns $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$, which can be given by

$$\begin{aligned} &\boldsymbol{m} = a_1\boldsymbol{m}_1 + a_2\boldsymbol{m}_2 \Leftrightarrow \\ &m_{n_y,n_z} = a_1 m(\mathbf{r}_{n_y,n_z}, \theta_1, \varphi_1) + a_2 m(\mathbf{r}_{n_y,n_z}, \theta_2, \varphi_2). \end{aligned} \tag{7}$$

Denote the signals transmitted to users as $\mathbf{s} = [s_1, s_2]^T$, and the precoding vector at the BS as $\mathbf{V} \in \mathbb{C}^{K\times 2}$. Since the RHS is a leaky-wave antenna, the leaky-wave effect makes the amplitude of the reference wave carrying the signals gradually decrease as it propagates along the waveguide [25]. Denote the attenuation constant during the propagation process of the reference wave as α, and thus, the amplitude attenuation of the reference wave when it propagates to the (n_y, n_z)-th element from feed k is $e^{-\alpha|\mathbf{r}^k_{n_y,n_z}|}$. Therefore, in HDMA, s_1 and s_2 are superposed as $\mathbf{x} = \mathbf{MVs}$, where $\mathbf{M} \in \mathbb{C}^{N_yN_z\times K}$ is composed of the following elements:

$$M^k_{n_y,n_z} = m_{n_y,n_z} \cdot e^{-\alpha|\mathbf{r}^k_{n_y,n_z}|} \cdot e^{-j\mathbf{k}_s\cdot\mathbf{r}^k_{n_y,n_z}}, \tag{8}$$

where $e^{-j\mathbf{k}_s\cdot\mathbf{r}^k_{n_y,n_z}}$ denotes the phase of the reference wave when it propagates to the (n_y, n_z)-th radiation element from the feed k. The received signal at user l can then be given by

$$y_l = \mathbf{H}_l\mathbf{x} + n_l = \mathbf{H}_l\mathbf{MVs} + n_l, \tag{9}$$

where $\mathbf{H}_l \in \mathbb{C}^{1\times N_yN_z}$ is the channel coefficient matrix between user l and the RHS, $n_l \sim \mathcal{CN}(\mathbf{0}, \sigma^2)$ is the additive white Gaussian noise (AWGN). The data rate of user l can then be given by

$$R_l = \log_2\left(1 + \frac{|\mathbf{H}_l\mathbf{MV}_l|^2}{\sigma^2 + \sum_{l'\neq l}|\mathbf{H}_l\mathbf{MV}_{l'}|^2}\right), \tag{10}$$

where $\mathbf{V}_l$ is the l-th column of $\mathbf{V}$.

In the HDMA, the RHS can generate two beams pointing to these two users through the superposed holographic pattern $\boldsymbol{m}$. For each user l, the interference from the other user can be alleviated by the precoder $\mathbf{V}$ at the BS such as the ZF precoder and the MMSE precoder [26], and thus, the quality-of-service of each user can be guaranteed. After user l receives the signal y_l, it down-converts the received signal to the baseband and then recovers the final signal.

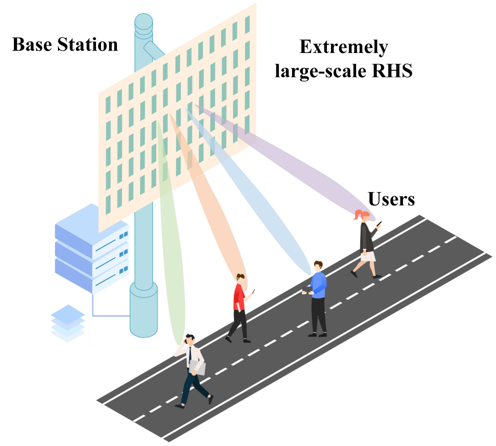

Fig. 4. HDMA wireless communication system aided by an extremely large-scale RHS

III. System Model

In this section, we introduce an HDMA wireless communication system where a BS with an extremely large-scale RHS serves multiple users via holographic beamforming, based on which the channel model is constructed.

[1]The specific transmission model for general HDMA wireless communication systems will be presented in Section III.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2022.3143221, IEEE Journal on Selected Areas in Communications

A. Scenario Description

As shown in Fig. 4, we consider a downlink HDMA wireless communication system, where a BS equipped with an extremely large-scale RHS transmits to multiple single-antenna users denoted as $\mathcal{L} = \{1, 2, \cdots, L\}$. By leveraging the principle of HDMA, the RHS maps all transmitted data to a single holographic pattern. When the reference wave carrying the transmitted signals excites the holographic pattern, the RHS can achieve beamforming without relying on complex phase-shifting circuits and bulky mechanics, and such a beamforming technique is also known as holographic beamforming.

B. Channel Model

Note that for an extremely large-scale RHS satisfying $N_y \to \infty$ and $N_z \to \infty$, the distance between the RHS and the user will be shorter than the Rayleigh distance, where $r_{Ray} = \frac{2D^2}{\lambda}$, $D = \max(N_y d_y, N_z d_z)$ is the maximum dimension of the RHS and λ is the wavelength [27]. Thus, the users are likely to be located in the near-field region of the RHS. The signal propagation distance and its phase are different with respect to each element, and thus, the conventional UPW model in the far-field region does not hold. The general spherical wave model is utilized to describe the variation of the signal propagation distances and its phase for different elements [28].

Denote the distance between each user l and the origin as d_l, the zenith and azimuth angles as θ_l and φ_l, and thus, the position vector of user l is $\mathbf{d}_l = [d_l \sin\theta_l \cos\varphi_l, d_l \sin\theta_l \sin\varphi_l, d_l \cos\theta_l]^T$. The distance vector from the (n_y, n_z)-th element to user l can then be given by $\mathbf{d}^l_{n_y,n_z} = \mathbf{d}_l - \mathbf{r}_{n_y,n_z}$. We adopt the free-space line-of-sight (LoS) propagation[2] in the considered HDMA system, and the reasons are given as below. For one thing, the LoS propagation is widely utilized in the theoretical analysis of the fundamental performance limits and asymptotic behaviors [29], [30], which can make the calculation of analytic bounds on system performance tractable. For the other, an extremely large-scale RHS is naturally deployed much higher above the sea level such as on the top of BSs, and thus, LoS paths will dominate.

Denote the size of each radiation element as $A = \sqrt{A} \times \sqrt{A}$, and thus, the surface domain of the (n_y, n_z)-th element is $A_{n_y,n_z} = [n_y d_y - \frac{\sqrt{A}}{2}, n_y d_y + \frac{\sqrt{A}}{2}] \times [n_z d_z - \frac{\sqrt{A}}{2}, n_z d_z + \frac{\sqrt{A}}{2}]$. Since the size of each radiation element is far smaller than the signal propagation distance between an element and the user, the variation of the signal propagation distance across different points in a radiation element is negligible. The channel power gain between user l and the (n_y, n_z)-th element can then be expressed as

$$|h^l_{n_y,n_z}|^2 = \iint_{A_{n_y,n_z}} G_l \left(\frac{\lambda}{4\pi |\mathbf{d}^l_{n_y,n_z}|} \right)^2 dydz = \frac{A\lambda^2 G_l}{(4\pi)^2 |\mathbf{d}^l_{n_y,n_z}|^2}, \tag{11}$$

[2]Since the free-space LoS propagation is adopted and the large-scale fading channel does not outdate in a period of time, the BS only needs to obtain the distance information. The exchange process of the signaling information between the BS and users for channel estimation is that each user sends pilot signals to the BS. The BS can then obtain the distance information based on the received signal strength (RSS) [31].

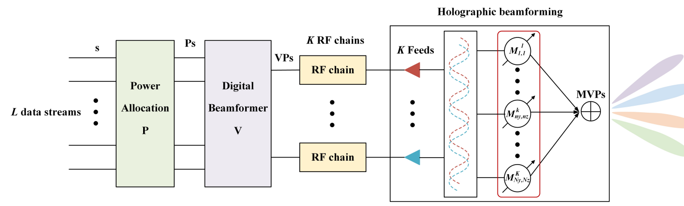

Fig. 5. HDMA transmission block diagram

where λ is the wavelength of the signal in free space, G_l is the antenna gain of each user. Therefore, considering the variation of signal phase for different elements, the channel coefficient between user l and the (n_y, n_z)-th element can be given by

$$\begin{aligned} h^l_{n_y,n_z} &= \frac{\lambda\sqrt{G_l}\sqrt{A} \cdot e^{-j\mathbf{k}_f \cdot \mathbf{d}^l_{n_y,n_z}}}{4\pi \|\mathbf{d}^l_{n_y,n_z}\|} \\ &= \frac{\lambda\sqrt{G_l}\sqrt{A} \cdot e^{-j\mathbf{k}_f \cdot \mathbf{d}^l_{n_y,n_z}}}{4\pi\sqrt{d_l^2 - 2d_l\Phi_l n_y d_y - 2d_l\Theta_l n_z d_z + n_y^2 d_y^2 + n_z^2 d_z^2}}, \end{aligned} \tag{12}$$

where $\mathbf{k}_f$ is the propagation vector in free space, $\Phi_l = \sin\theta_l \sin\varphi_l$, and $\Theta_l = \cos\theta_l$.

IV. Multi-User Holographic Beamforming

In this section, we present the multi-user holographic beamforming scheme for HDMA. Following that, the capacity of the HDMA wireless communication system is given.

To serve the users simultaneously, as shown in Fig. 5, the BS first encodes the data streams for different users via a digital beamformer $\mathbf{V} \in \mathbb{C}^{K \times L}$ at baseband, since the RHS does not have any digital processing capability. The processed signals are then up-converted to the carrier frequency by passing through K RF chains. Specifically, each RF chain is connected with a feed of the RHS. After up-converting the transmitted signals to the carrier frequency, each RF chain sends the up-converted signals to its connected feed. The feed then transforms the high-frequency current into the electromagnetic wave, which is also called reference wave, propagating on the RHS. The reference waves will be transformed into leaky waves through radiation elements on the RHS and leak out energy into free space for radiation, where the radiation amplitude of the reference wave at each radiation element is controlled via a holographic beamformer $\mathbf{M} \in \mathbb{C}^{N_y N_z \times K}$ to generate desired directional beams pointing to the users. Based on the principle of HDMA, the holographic beamforming matrix $\mathbf{M} \in \mathbb{C}^{N_y N_z \times K}$ is then formed by the following elements

$$M^k_{n_y,n_z} = m_{n_y,n_z} \cdot e^{-\alpha|\mathbf{r}^k_{n_y,n_z}|} \cdot e^{-j\mathbf{k}_s \cdot \mathbf{r}^k_{n_y,n_z}}, \tag{13}$$

where m_{n_y,n_z} is the holographic pattern given in (6), and $e^{-j\mathbf{k}_s \cdot \mathbf{r}^k_{n_y,n_z}}$ denotes the phase of the reference wave when it propagates to the (n_y, n_z)-th radiation element from feed k.

Without loss of generality, it is assumed that the signals are transmitted under a power allocation matrix[3] $\mathbf{P}$ satisfying

[3]Since we mainly focus on the holographic pattern design for asymptotic capacity analysis, the power allocation matrix $\mathbf{P}$ is assumed to be a constant matrix in the following.

$$[\mathbf{G}]_{l,k} = \frac{\lambda\sqrt{G_l}\sqrt{A}d_l e^{-j\mathbf{k}_f\cdot\mathbf{d}_l}}{4\pi d_y d_z}\iint_{\mathcal{A}} \frac{e^{-\alpha d_l\sqrt{(y-y_k/d_l)^2+(z-z_k/d_l)^2}-j\omega_{l,k}}}{\sqrt{1-2y\Phi_l-2z\Theta_l+y^2+z^2}}\cdot\left[\sum_{l'=1}^{L}\sum_{k'=1}^{K}\frac{a_{l',k'}(1+\cos\omega_{n_y,n_z}^{l',k'})}{2}\right]dydz. \tag{17}$$

$\mathrm{Tr}\{\mathbf{PP}^H\} = P$, where P is the total transmit power at the BS. Denote the intended signal vector for L users as $\mathbf{s} \in \mathbb{C}^{L\times 1}$ satisfying $\mathbb{E}[\mathbf{ss}^H] = \mathbf{I}_L$. Therefore, the transmitted signals at the BS are $\mathbf{VPs}$. After the holographic beamforming by the RHS integrated with the BS, the received signal at the users can then be given by

$$\mathbf{y} = \mathbf{HMVPs} + \mathbf{n}, \tag{14}$$

where $\mathbf{H} \in \mathbb{C}^{L\times N_y N_z}$ is the channel matrix composed of element h_{n_y,n_z}^l as given in (12), $\mathbf{M}$ is the holographic beamforming matrix as given in (13), and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0},\sigma^2)$ is the AWGN.

The capacity of the HDMA wireless communication system can be achieved by dirty paper coding (DPC) [32], where the interference among the users can be eliminated by utilizing the DPC scheme at the BS. Specifically, the capacity can be given by [33]

$$C = \sum_{l=1}^{L}\log_2\left(1+\frac{P_l}{\sigma^2}[\mathbf{GG}^H]_{l,l}\right), \tag{15}$$

where $P_l = [\mathbf{PP}^H]_{l,l}$, the equivalent channel matrix $\mathbf{G} = \mathbf{HM}$ consists of following elements:

$$[\mathbf{G}]_{l,k} = \sum_{n_y=-\frac{N_y-1}{2}}^{\frac{N_y-1}{2}}\sum_{n_z=-\frac{N_z-1}{2}}^{\frac{N_z-1}{2}} h_{n_y,n_z}^l M_{n_y,n_z}^k. \tag{16}$$

Based on the channel model given in (12) and the holographic beamforming matrix given in (13), the following proposition about the specific expression of $[\mathbf{G}]_{l,k}$ can be derived.

Proposition 1. The equivalent channel $[\mathbf{G}]_{l,k}$ can be expressed as (17), where $\mathcal{A} = \left\{(y,z)\middle| |y| \le \frac{N_y d_y}{2d_l}, |z| \le \frac{N_z d_z}{2d_l}\right\}$ is the integral domain, (y_k, z_k) is the position of feed k, and

$$\begin{aligned}\omega_{n_y,n_z}^{l,k} &= \mathbf{k}_f\cdot\mathbf{r}_{n_y,n_z} - \mathbf{k}_s\cdot\mathbf{r}_{n_y,n_z}^k\\ &= d_l\cdot|\mathbf{k}_f|\cdot(y\Phi_l+z\Theta_l) -\\ &\quad d_l\cdot|\mathbf{k}_\mathbf{s}|\cdot\sqrt{(y-y_k/d_l)^2+(z-z_k/d_l)^2}.\end{aligned} \tag{18}$$

Proof: See Appendix A. ∎

V. Holographic Pattern Design

In this section, we develop a holographic pattern optimization scheme to achieve the asymptotic capacity of the HDMA wireless communication system aided by an extremely large-scale RHS.

Based on the expression of the capacity given in (15), the holographic pattern optimization problem can be formulated as

$$\max_{\{a_{l,k}\}}\sum_{l=1}^{L}\log_2\left(1+\frac{P_l}{\sigma^2}[\mathbf{GG}^H]_{l,l}\right), \tag{19a}$$

$$s.t. \quad \sum_{l=1}^{L}\sum_{k=1}^{K}a_{l,k} = 1. \tag{19b}$$

Remark 1. According to the above holographic pattern optimization problem, it can be seen that the number of variables to be optimized (i.e., $\{a_{l,k}\}$) is $L\times K$. Differently, in a traditional SDMA system with K RF chains and N transmit antennas, the size of the phase-controlled analog beamformer is $N\times K$, indicating that $N\times K$ phase shifters need to be optimized [10]. Since the number of users is less than the number of antenna elements, i.e., $L < N$ in general, the precoding complexity of the HDMA system is less than that of the traditional SDMA system. Besides, different from the traditional SDMA system relying on complex phase-shifting circuits, HDMA with RHS is realized by metamaterial radiation elements with controllable radiation amplitude. Such metamaterial radiation elements can be fabricated by loading low-power and low-complexity RF switches such as PIN diodes, and their radiation amplitude can be changed by controlling the biased voltage applied to these RF switches [22]. Thus, the HDMA system is envisioned to be of lower complexity than the traditional SDMA system.

To solve this problem, we first present the following two lemmas about the properties of the equivalent channel matrix $\mathbf{G}$ when the RHS is extremely large.

Lemma 1. For an extremely large-scale RHS satisfying $N_y \to \infty$ and $N_z \to \infty$, $[\mathbf{G}]_{l,k}$ can be approximated by

$$\begin{aligned}[\mathbf{G}]_{l,k} =&\frac{a_{l,k}\lambda\sqrt{G_l}\sqrt{A}d_l\cdot e^{-j\mathbf{k}_f\cdot\mathbf{d}_l}}{16\pi d_y d_z}\cdot\\ &\iint_{\mathcal{A}}\frac{e^{-\alpha d_l\sqrt{(y-y_k/d_l)^2+(z-z_k/d_l)^2}}}{\sqrt{1-2y\Phi_l-2z\Theta_l+y^2+z^2}}dydz.\end{aligned} \tag{20}$$

Proof: See Appendix B. ∎

The following lemma can then be derived from Lemma 1.

Lemma 2. For an extremely large-scale RHS satisfying $N_y \to \infty$ and $N_z \to \infty$, $[\mathbf{GG}^H]_{l,l}$ can be given by

$$[\mathbf{GG}^H]_{l,l} = \frac{\lambda^2 G_l A d_l^2}{256\pi^2 d_y^2 d_z^2}\sum_{k=1}^{K}a_{l,k}^2 I_{l,k}^2, \tag{21}$$

where $I_{l,k}$ is defined as

$$I_{l,k} = \iint_{\mathcal{A}}\frac{e^{-\alpha d_l\sqrt{(y-y_k/d_l)^2+(z-z_k/d_l)^2}}}{\sqrt{1-2y\Phi_l-2z\Theta_l+y^2+z^2}}dydz. \tag{22}$$

The holographic pattern optimization problem given in (19) can then be rewritten as

$$\max_{\{a_{l,k}\}}\sum_{l=1}^{L}\log_2\left(1+\frac{P_l\lambda^2 G_l A d_l^2}{256\sigma^2\pi^2 d_y^2 d_z^2}\sum_{k=1}^{K}a_{l,k}^2 I_{l,k}^2\right), \tag{23a}$$

$$s.t. \quad \sum_{l=1}^{L}\sum_{k=1}^{K}a_{l,k} = 1. \tag{23b}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2022.3143221, IEEE Journal on Selected Areas in Communications

Proposition 2. The capacity can be upper bounded by

$$C \leq \sum_{l=1}^{L} \log_2 \left(1 + \frac{P_l \lambda^2 G_l A d_l^2}{256\sigma^2\pi^2 d_y^2 d_z^2} (\sum_{k=1}^{K} a_{l,k})^2 (\max_{1\leq k\leq K} I_{l,k})^2 \right), \tag{24}$$

where the equality holds when $a_{l,k}$ satisfies[4]

$$\begin{cases} a_{l,k} \neq 0, & k = k_l^*, \\ a_{l,k} = 0, & k \neq k_l^*, \end{cases} \tag{25}$$

where for each user l, $k_l^* = \arg\max_{1\leq k\leq K} I_{l,k}$.

Proof: See Appendix C. ∎

Remark 2. Proposition 2 reveals that in an HDMA wireless communication system, the asymptotic capacity can be achieved if and only if the holographic pattern $\boldsymbol{m}$ is formed by the following elements

$$m_{n_y,n_z} = \sum_{l=1}^{L} a_l^* m(\mathbf{r}_{n_y,n_z}^{k_l^*}, \theta_l, \varphi_l), \tag{26}$$

where a_l^* will be given in (29), $m(\mathbf{r}_{n_y,n_z}^{k_l^*}, \theta_l, \varphi_l)$ is given in (5).

For notational convenience, we define $I_l = \frac{P_l\lambda^2 G_l A d_l^2}{256\sigma^2\pi^2 d_y^2 d_z^2}(I_{l,k_l^*})^2$ and $a_l = a_{l,k_l^*}$. Therefore, based on (23), the holographic pattern optimization problem can be simplified as

$$\max_{\{a_l\}} \sum_{l=1}^{L} \log_2\left(1 + I_l a_l^2\right), \tag{27a}$$

$$s.t. \quad \sum_{l=1}^{L} a_l = 1. \tag{27b}$$

To derive the optimal solution of $\{a_l^*\}$, we adopt the Lagrangian dual form to relax the constraint (27b) with a multiplier since problem (27) is a maximization problem with an equality constraint. Denote β as the Lagrangian multiplier associated with the constraint (27b). The Lagrangian associated with holographic pattern design can be given by

$$\boldsymbol{\mathcal{L}}(a_l, \beta) = \sum_{l=1}^{L} \log_2\left(1 + I_l a_l^2\right) - \beta(\sum_{l=1}^{L} a_l - 1). \tag{28}$$

By setting $\partial\boldsymbol{\mathcal{L}}/\partial a_l = 0$, the optimal $\{a_l^*\}$ and β^* can be obtained by solving the following system of equations:

$$\begin{cases} a_l^* = \dfrac{1}{\beta^* \ln 2} + \sqrt{\dfrac{1}{(\beta^*\ln 2)^2} - \dfrac{1}{I_l}}, \\ \displaystyle\sum_{l=1}^{L} a_l^* = 1. \end{cases} \tag{29}$$

Therefore, the holographic pattern $\boldsymbol{m}$ can be obtained by substituting $\{a_l^*\}$ into (6), i.e., the holographic pattern $\boldsymbol{m}$ is formed by the following elements

$$m_{n_y,n_z} = \sum_{l=1}^{L} a_l^* m(\mathbf{r}_{n_y,n_z}^{k_l^*}, \theta_l, \varphi_l), \tag{30}$$

[4] For simplicity, we assume that for each fixed l, there exists a unique k maximizing $\{I_{l,k}\}_{k=1}^{K}$.

where $m(\mathbf{r}_{n_y,n_z}^{k_l^*}, \theta_l, \varphi_l)$ is given in (5).

Proposition 3. When I_l satisfies $\min_l I_l \geq 12(L-1)^2$, the optimal solution given by (29) is the global optimal solution to problem (19)[5].

Proof: See Appendix D. ∎

Based on (29) and Proposition 2, we have the following proposition.

Proposition 4. For an extremely large-scale RHS satisfying $N_y \to \infty$ and $N_z \to \infty$, the capacity of the HDMA wireless communication system can be given by

$$\begin{aligned} \widetilde{C} &= \sum_{l=1}^{L} \log_2\left(1 + I_l (a_l^*)^2\right) \\ &= \sum_{l=1}^{L} \log_2\left(1 + \frac{P_l\lambda^2 G_l A d_l^2}{256\sigma^2\pi^2 d_y^2 d_z^2}(a_l^*)^2 (I_{l,k_l^*})^2\right), \end{aligned} \tag{31}$$

where a_l^* is given in (29) and I_{l,k_l^*} can be obtained by (22).

Remark 3. The assumption of an extremely large-scale RHS is only utilized in the proof of Lemma 1 (i.e., Appendix B) to make the proof more rigorous mathematically. In Section VII, we will verify that the holographic pattern design scheme and performance analysis in the HDMA system with an extremely large-scale RHS also hold for the HDMA system with a normal-sized RHS.

VI. Performance Analysis

In this section, we first analyze the relation between the sum rate with ZF precoding and the capacity of the HDMA wireless communication system. It is demonstrated that for an extremely large-scale RHS, the optimal holographic pattern proposed in Section V can make the sum rate with ZF precoding achieve the asymptotic capacity. Following that, we consider a special case of a single-user communication scenario where the specific forms of upper and lower bounds of the data rate are given. Finally, we compare the performance of the HDMA system and the traditional SDMA system.

A. Sum Rate

Since the ZF precoder can obtain a near-optimal solution with low complexity, we consider ZF precoding together with power allocation at the BS to alleviate the inter-user interference [8]. The l-th column of the ZF precoder $\mathbf{V} \in \mathbb{C}^{K\times L}$ can be given by

$$\mathbf{V}_l = \frac{\mathbf{W}_l}{|\mathbf{W}_l|}, \tag{32}$$

where $\mathbf{W}_l$ is the l-th column of matrix $\mathbf{G}^H(\mathbf{G}\mathbf{G}^H)^{-1}$ and $\mathbf{G} = \mathbf{H}\mathbf{M}$. Based on the expression of $\mathbf{V}$, the following proposition can be derived.

Proposition 5. The data rate of user l can be given by

$$R_l = \log_2\left(1 + \frac{P_l}{\sigma^2[(\mathbf{G}\mathbf{G}^H)^{-1}]_{l,l}}\right). \tag{33}$$

Proof: See Appendix E. ∎

[5] When $N_y = N_z = 100$, under the parameter settings given in Section VII, the magnitude of $\min_l I_l$ is about 10^5. Therefore, for an extremely large-scale RHS satisfying $N_y \to \infty$ and $N_z \to \infty$, the condition $\min_l I_l \geq 12(L-1)^2$ can be easily satisfied.

$$k_l^* = \arg\min_{1\le k\le K} d_k^l = \arg\min_{1\le k\le K} \sqrt{(y_k - d_l\Phi_l)^2 + (z_k - d_l\Theta_l)^2 + (d_l\sin\theta_l\cos\phi_l)^2}. \tag{35}$$

$$R_s = \log_2\left(1+\frac{P|\mathbf{H}\mathbf{M}|^2}{\sigma^2}\right) = \log_2\left(1+\frac{P\lambda^2 GA}{64\sigma^2\pi^2}\left|\sum_{n_y=-\frac{N_y-1}{2}}^{\frac{N_y-1}{2}}\sum_{n_z=-\frac{N_z-1}{2}}^{\frac{N_z-1}{2}}\frac{e^{-\alpha\sqrt{n_y^2d_y^2+n_z^2d_z^2}}\cdot e^{-j\omega_{n_y,n_z}}\cdot(1+\cos\omega_{n_y,n_z})}{\sqrt{d^2-2d\Phi n_y d_y - 2d\Theta n_z d_z + n_y^2d_y^2+n_z^2d_z^2}}\right|^2\right). \tag{39}$$

Therefore, the sum rate with ZF precoding of the HDMA communication system can be given by

$$R = \sum_{l=1}^{L} R_l = \sum_{l=1}^{L}\log_2\left(1+\frac{P_l}{\sigma^2[(\mathbf{G}\mathbf{G}^H)^{-1}]_{l,l}}\right). \tag{34}$$

B. Relation between Sum Rate and Capacity

Based on Proposition 2, the asymptotic capacity is determined by $\{a_{l,k}\}$, which is related to k_l^* and $I_{l,k}$. The properties about k_l^* and $I_{l,k}$ are given in the following lemma.

Lemma 3. Define the distance between user l and feed k as d_k^l. For an extremely large-scale RHS satisfying $N_y \to \infty$ and $N_z \to \infty$, $I_{l,k}$ given in (22) decreases as d_k^l grows, and thus, k_l^* can be given by (35).

Proof: See Appendix F. ■

Combining Proposition 2 and Lemma 3 also gives the following proposition about the relation between the sum rate with ZF precoding and the asymptotic capacity of the HDMA wireless communication system.

Proposition 6. In an HDMA wireless communication system aided by an extremely large-scale RHS, the sum rate with ZF precoding can achieve the asymptotic capacity when the following conditions are satisfied: (1) The feed closest to each user is different, i.e., $k_{l_1}^* \neq k_{l_2}^*, \forall l_1 \neq l_2$; (2) The holographic pattern $\boldsymbol{m}$ is formed by the following elements

$$m_{n_y,n_z} = \sum_{l=1}^{L} a_l^* m(\mathbf{r}_{n_y,n_z}^{k_l^*}, \theta_l, \varphi_l), \tag{36}$$

where a_l^* is given in (29), $m(\mathbf{r}_{n_y,n_z}^{k_l^*}, \theta_l, \varphi_l)$ is given in (5).

Proof: See Appendix G. ■

C. Special Case: Data Rate of the Single-User Communication System

In the single-user wireless communication case where the BS with an RHS transmits to a single user, the received signal at the user can be given by

$$y = \mathbf{H}\mathbf{M}\sqrt{P}s + w = \sum_{n_y=-\frac{N_y-1}{2}}^{\frac{N_y-1}{2}}\sum_{n_z=-\frac{N_z-1}{2}}^{\frac{N_z-1}{2}} h_{n_y,n_z}M_{n_y,n_z}\sqrt{P}s + w, \tag{37}$$

where the user index l is omitted, h_{n_y,n_z} is given in (12). Based on (5) and (13), M_{n_y,n_z} can be given by

$$M_{n_y,n_z} = \frac{\cos\omega_{n_y,n_z}+1}{2}\cdot e^{-\alpha|\mathbf{r}_{n_y,n_z}|}\cdot e^{-j\mathbf{k}_s\cdot\mathbf{r}_{n_y,n_z}}, \tag{38}$$

where $\omega_{n_y,n_z} = \mathbf{k}_f\cdot\mathbf{r}_{n_y,n_z} - \mathbf{k}_s\cdot\mathbf{r}_{n_y,n_z}$.

Therefore, based on (37) and (38), the data rate in the single-user case can be given by (39).

Similar to the multi-user case, according to Proposition 2, for an extremely large-scale RHS satisfying $N_y \to \infty$ and $N_z \to \infty$, R_s can be approximated by

$$R_s \approx \log_2\left(1+\frac{P\lambda^2 GAd^2}{256\sigma^2\pi^2 d_y^2 d_z^2} I^2\right), \tag{40}$$

where I is defined as

$$I = \iint_{\mathcal{A}} \frac{e^{-\alpha d\sqrt{y^2+z^2}}}{\sqrt{1-2y\Phi-2z\Theta+y^2+z^2}} dydz. \tag{41}$$

To derive the upper and lower bounds of R_s, we present the following lemma about the region of I.

Lemma 4. The region of I can be given by (42), which is shown in the top of the next page.

Proof: See Appendix H. ■

Proposition 7. For an extremely large-scale RHS satisfying $N_y \to \infty$ and $N_z \to \infty$, the upper bound of the data rate in the single-user case can be given by

$$R_s \le \log_2\left(1+\frac{P\lambda^2GAd^2}{256\sigma^2d_y^2d_z^2}\left[e^{-\alpha d\Omega}(I_1+\Omega I_2) + \frac{1-(\alpha d\Omega+1)e^{-\alpha d\Omega}}{\alpha^2d^2\Psi} + \frac{4-\pi\alpha d\mathrm{Y}_1(\alpha d)-\pi\alpha d\mathrm{H}_{-1}(\alpha d)}{2\alpha d}\right]^2\right), \tag{43}$$

where $\Omega = \sqrt{\Phi^2+\Theta^2}$, $\Psi = \sin\theta\cos\varphi$, $\mathrm{Y}_\nu(\cdot)$ is the Bessel function of the second kind, $\mathrm{H}_\nu(\cdot)$ is the Struve function [34]. I_1 and I_2 can be expressed as

$$\begin{aligned} I_1 &= \frac{4-\pi\alpha d\Psi\mathrm{Y}_1(\alpha d\Psi)-\pi\alpha d\Psi\mathrm{H}_{-1}(\alpha d\Psi)}{2\alpha d}, \\ I_2 &= \frac{\pi}{2}[\mathrm{H}_0(\alpha d\Psi)-\mathrm{Y}_0(\alpha d\Psi)]. \end{aligned} \tag{44}$$

Proof: See Appendix I. ■

Proposition 8. For an extremely large-scale RHS satisfying $N_y \to \infty$ and $N_z \to \infty$, the lower bound of the data rate in the single-user case can be given by

$$R_s \ge \log_2\left\{1+\frac{P\lambda^2GAd^2}{256\sigma^2d_y^2d_z^2}\cdot \left[\frac{4-\pi\alpha d\mathrm{Y}_1(\alpha d)-\pi\alpha d\mathrm{H}_{-1}(\alpha d)}{\alpha d}\right]^2\right\}. \tag{45}$$

Proof: See Appendix I. ■

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2022.3143221, IEEE Journal on Selected Areas in Communications

$$
\iint_{\mathcal{A}} \frac{e^{-\alpha d\sqrt{y^2+z^2}}}{\sqrt{1+y^2+z^2}} dydz \le I \le \frac{1}{2}\left[\iint_{\mathcal{A}} \frac{e^{-\alpha d\sqrt{y^2+z^2}}}{\sqrt{1-2\sqrt{(y^2+z^2)(\Phi^2+\Theta^2)}+y^2+z^2}} dydz + \iint_{\mathcal{A}} \frac{e^{-\alpha d\sqrt{y^2+z^2}}}{\sqrt{1+2\sqrt{(y^2+z^2)(\Phi^2+\Theta^2)}+y^2+z^2}} dydz\right]. \tag{42}
$$

D. Comparison of the HDMA System and the Traditional SDMA System

In this subsection, we compare the performance of the HDMA system and that of the traditional SDMA system. Since the element spacing and the number of radiation elements are closely related to the capacity, we first give the following proposition about the influence of the element spacing and the number of radiation elements on the capacity.

Proposition 9. In the HDMA system, when the physical dimension of the RHS along the y-axis and z-axis are fixed, i.e., $N_y d_y$ and $N_z d_z$ are constants, the capacity increases logarithmically as the inverse of the element spacing d_y and d_z grows, i.e., $C \propto \log_2 (d_y d_z)^{-1}$. The capacity also increases logarithmically with the element number N, i.e., $C \propto \log_2 N$, where $N = N_y N_z$.

Proof: See Appendix J. ■

In the SDMA system, when the physical dimension of the phased array is fixed, the capacity also increases logarithmically with the element number[6] N, i.e., $C \propto \log_2 N$ [9]. Considering that when the physical dimensions of an RHS and a phased array are the same, the number of elements in the RHS is much larger than that of the phased array due to the compact element spacing of the RHS. Therefore, the HDMA system has great potential in enhancing the capacity and supporting massive connectivity, which will be verified in Section VII.

Remark 4. Since the reference wave leaks out energy at each element during its propagation process, the remaining power of the reference wave at each element is influenced by the radiated energy in previous elements along its propagation path. Therefore, the transmit power is equal to the sum of the total radiated power from the RHS and the power lost during the wave propagation, i.e., $e^{-\alpha|\mathbf{r}^k_{n_y,n_z}|}$.

However, the total radiated power from the RHS is hard to be depicted due to the coupling relation between the adjacent elements. Therefore, we consider an effective $\hat{\alpha}$ such that $e^{-\hat{\alpha}|\mathbf{r}^k_{n_y,n_z}|}$ can guarantee the sum of the total radiated power from the RHS and the power lost during the wave propagation is no larger than the transmit power[7]. In this case, since the transmit power is not fully utilized, (31) presents the lower bound of the HDMA system capacity when adopting the effective $\hat{\alpha}$. Moreover, we will also verify that the decrease of the capacity brought by $\hat{\alpha}$ can be compensated by the compact element spacing in the RHS in Section VII.

[6]The proof of this statement will also be given in Appendix J.

[7]The value of $\hat{\alpha}$ can be determined by numerical simulations varying with the physical dimension of the RHS.

TABLE I
SIMULATION PARAMETERS

Parameters	Values
Transmit power of the BS P (W)	10
Element spacing of the RHS d_y and d_z (cm)	0.75
Carrier frequency f (GHz)	10
Propagation vector in the free space k_f	$200\pi/3$
Propagation vector on the RHS k_s	$200\sqrt{3}\pi/3$
Effective attenuation constant $\hat{\alpha}$	8
Number of users L	5
Antenna gain of each user G_l (dBi)	10
Noise density σ_l^2 (dBm/Hz)	-174

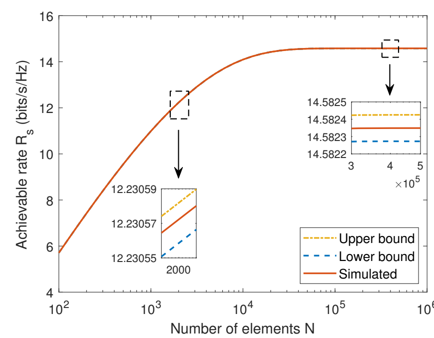

Fig. 6. Data rate in the single-user case versus N.

VII. SIMULATION RESULTS

In this section, we evaluate the performance of the HDMA wireless communication system to validate the theoretical analysis. Without loss of generality, it is assumed that the zenith angles of the users $\{\theta_l\}$ are uniformly distributed in $[0^\circ, 90^\circ]$, while the azimuth angles of the users $\{\varphi_l\}$ are uniformly distributed in $[-90^\circ, 90^\circ]$, and the distances between the users and the origin are uniformly distributed in $[50\text{m}, 100\text{m}]$. The number of the feeds of the RHS is set as the same as the number of users [16]. Major simulation parameters are set up based on the existing works [33], [35] and 3GPP specifications [36] as given in Table I. Moreover, since we adopt an effective $\hat{\alpha}$, the simulation results present the lower bound of the capacity and the sum rate of the HDMA system.

Fig. 6 illustrates the data rate in the special single-user case R_s versus the number of elements in the RHS N. It can be seen that R_s increases with N and converges to a constant value when N is extremely large. More importantly, Fig. 6 also shows that our closed-form upper and lower rate bounds depicted by (43) and (45) give good approximations of R_s.

Fig. 7 compares the cost-efficiency of the HDMA scheme and traditional SDMA scheme, where the physical dimension of the RHS and that of the phased array are the same and fixed. Specifically, the cost-efficiency metric η is defined as the ratio of sum rate to hardware cost. In general, hardware cost ratio of phased array to the RHS of the same number of elements

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2022.3143221, IEEE Journal on Selected Areas in Communications

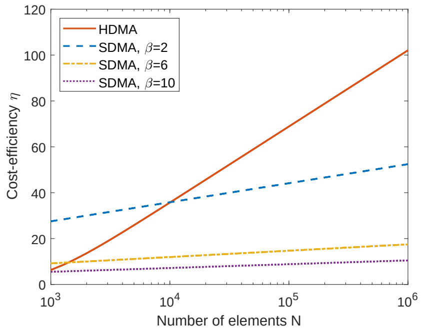

Fig. 7. Cost-efficiency versus the number of elements N.

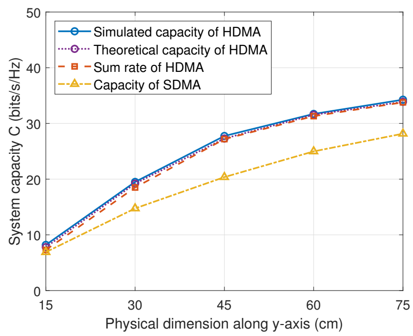

Fig. 8. System capacity versus the physical dimension.

β is about $2 \sim 10$ [22], since the phased array requires high-priced electronic components such as phase shifters at each element. It can be seen that the cost-efficiency of the HDMA scheme is first lower and then greater than that of the SDMA scheme. The main reason is that $\hat{\alpha}$ decreases the sum rate in the HDMA scheme, and thus, the phased array can achieve a higher sum-rate than the RHS with the same number of elements. However, as the number of elements grows, the advantage in achieving higher sum rate of the phased array is insufficient to offset its high hardware cost. Moreover, the superiority of the HDMA scheme in cost savings becomes clearer with the number of elements and the cost ratio β.

Fig. 8 depicts the capacity of the HDMA wireless communication system C versus the physical dimension of the RHS, where the optimal holographic pattern given in (30) is adopted. We observe that not only for an extremely large-scale RHS, but also for a normal-sized RHS, the optimal holographic pattern can make the sum rate with ZF precoding of the HDMA system achieve the capacity C. The accuracy of the approximation of the equivalent channel $[\mathbf{G}]_{l,k}$ given in Lemma 1 is also verified.

Fig. 9 shows the maximum number of accessed users L in the HDMA wireless communication system with ZF precoding versus the physical dimension of the RHS. For each user, the threshold of the data rate is set as 5 bits/s/Hz, i.e., when the total bandwidth is 2MHz, the data rate of each user should be larger than 10Mbps. We observe that the maximum number of accessed users grows with the physical dimension of the RHS and the transmit power P, since the data rate of each user grows with the number of elements in the RHS and P.

Figs. 8 and 9 also compare the performance of an HDMA wireless communication system with a traditional multi-user MIMO system utilizing SDMA in terms of capacity and the

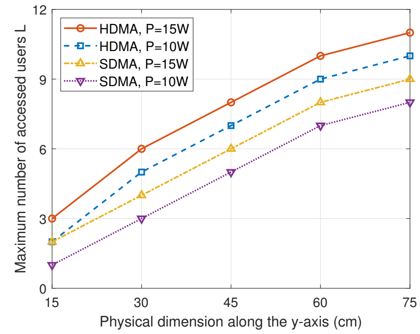

Fig. 9. Number of accessed users versus the physical dimension.

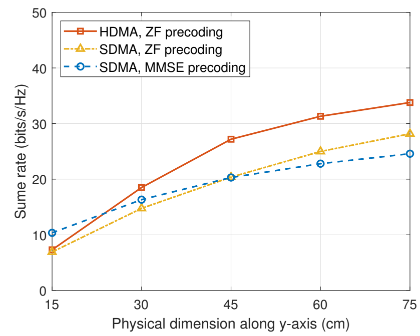

Fig. 10. Sum rate versus the physical dimension.

maximum number of accessed users. For the sake of fairness, we assume that the physical dimension of the phased array and that of the RHS are the same. The element spacing of the phased array in the multi-user MIMO system is set as half wavelength [10], and that of the RHS is set as one quarter wavelength according to the effective medium theory [13]. In the traditional multi-user MIMO system, the ZF precoding is performed, based on which the traditional analog beamformer is optimized by the coordinate ascent algorithm. It can be seen that the HDMA system outperforms the traditional multi-user MIMO system in terms of capacity and the maximum number of accessed users as physical dimension of the RHS grows. It can be seen that the HDMA system significantly outperforms the traditional multi-user MIMO system in terms of capacity and the maximum number of accessed users, since the compact element spacing leads to much more radiation elements in the RHS. This reveals the potential of the HDMA wireless communication system in achieving massive connectivity.

Considering that the ZF-based precoding scheme is not near-optimal at the low signal-to-noise ratio (SNR) regime in the traditional multi-user MIMO system, Fig. 10 further compares the sum rate of the HDMA wireless communication system with that of the traditional multi-user MIMO system, where both ZF and minimum mean square error (MMSE)-based solutions are considered for digital precoding in the MIMO system. It can be seen that the HDMA wireless communication system also outperforms the traditional multi-user MIMO system with ZF-based digital precoding or MMSE-based digital precoding in terms of the sum rate.

VIII. Conclusions

In this paper, we have presented the concept of HDMA for future wireless communications, where the holographic

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2022.3143221, IEEE Journal on Selected Areas in Communications

pattern construction of the RHS and the HDMA principles have been elaborated. Specifically, utilizing the superposition of the holographic patterns corresponding to the receivers, the intended signals can be mapped to a superposed holographic pattern. Based on the multi-user holographic beamforming scheme for HDMA, the asymptotic capacity and sum rate of an HDMA wireless communication system aided by an extremely large-scale RHS have been analyzed. The closed-form optimal holographic pattern to achieve asymptotic capacity has been derived. In addition, considering the special single-user case, we have derived the closed-form lower and upper bounds of the data rate, which are verified to give an accurate theoretical approximation of the data rate in the simulations.

Three conclusions can be drawn from the theoretical analysis and simulation results, providing insights for the HDMA wireless communication system design. *First*, based on the derived optimal holographic pattern, the sum rate with simple ZF precoding of the HDMA wireless communication system can achieve the capacity. Such a statement also holds for an HDMA wireless communication system aided by a normal-sized RHS. *Second*, with the same element spacing of the antenna array, the HDMA scheme provides a powerful solution to reduce the hardware cost while guaranteeing a high sum-rate compared with traditional SDMA scheme. *Third*, with a compact element spacing of the RHS, the HDMA wireless communication system outperforms the traditional multi-user MIMO system in terms of capacity and the maximum number of accessed users. This indicates that the HDMA scheme has great potential in capacity improvement and enhancing massive connectivity.

For future outlook, we remark that although HDMA has unique advantages over traditional SDMA, several key challenges remain to be solved such as the channel estimation scheme for an ultra-large RHS and the incorporation with other multiple access technologies. Specifically, for one thing, due to the compact element spacing of the RHS, the overhead required for channel estimation will be overwhelming for an ultra-large RHS. Thus, a fast and accurate channel estimation scheme coupled with the working principle of HDMA is urgently required to reduce pilot training overhead and channel state information feedback efficiently. For the other, HDMA exploiting the spatial domain can be incorporated with other multiple access technologies such as orthogonal frequency division multiple access (OFDMA) exploiting the frequency domain and non-orthogonal multiple access (NOMA) exploiting the power domain. Motivated by the characteristics of different multiple access technologies, the potential benefits brought by OFDMA-HDMA systems and NOMA-HDMA systems are worthy to be explored.

Appendix A
Proof of Proposition 1

Based on (12) and (13), the equivalent channel $[\mathbf{G}]_{l,k}$ can be given by (46). Define the function $g(y,z)$ as

$$g(y,z) = \frac{e^{-\alpha d_l\sqrt{(y-y_k/d_l)^2+(z-z_k/d_l)^2}}}{\sqrt{1-2\Phi_l y-2\Theta_l z+y^2+z^2}}\cdot e^{-j\omega^{k,l}_{n_y,n_z}}\cdot \left[\sum_{l'=1}^{L}\sum_{k'=1}^{K}\frac{a_{l',k'}(1+\cos\omega^{l',k'}_{n_y,n_z})}{2}\right]. \tag{47}$$

Since $d_y \ll d_l$, $d_z \ll d_l$, based on the definition of double integral, we have $\sum\limits_{n_y,n_z} g\left(\frac{n_y d_y}{d_l},\frac{n_z d_z}{d_l}\right) \approx \frac{d_l^2}{d_y d_z}\cdot \iint_{|y|\le\frac{N_y d_y}{2d_l},|z|\le\frac{N_z d_z}{2d_l}} g(y,z)dydz$, where $y=\frac{n_y d_y}{d_l}, z=\frac{n_z d_z}{d_l}$. $[\mathbf{G}]_{l,k}$ can then be expressed as (48), where $\mathcal{A}=\left\{(y,z)\middle| |y|\le\frac{N_y d_y}{2d_l}, |z|\le\frac{N_z d_z}{2d_l}\right\}$. This completes the proof.

Appendix B
Proof of Lemma 1

Note that $e^{-j\omega^{l,k}_{n_y,n_z}}\cdot\left[\sum_{l'=1}^{L}\sum_{k'=1}^{K}\frac{a_{l',k'}(1+\cos\omega^{l',k'}_{n_y,n_z})}{2}\right]$ can be rewritten as (49), which is shown in the top of the next page. We need to prove that for an extremely large-scale RHS satisfying $N_y\to\infty$ and $N_z\to\infty$, the integral of the items containing sine and cosine functions can be neglected, such that only the first item $\frac{a_{l,k}}{4}$ remains. For brevity, we only give proof for the item $\frac{\cos\omega^{l,k}_{n_y,n_z}}{2}$. The proof for the other items can be obtained similarly.

For an extremely large-scale RHS, utilizing the polar coordinate transformation $y-d_k/d_l=\rho\cos\vartheta,\ z-z_k/d_l=\rho\sin\vartheta$, the integral in (17) corresponding to the item $\frac{\cos\omega^{l,k}_{n_y,n_z}}{2}$ is

$$\begin{aligned}&\frac{1}{2}\iint_{\mathbb{R}^2}\frac{e^{-\alpha d_l\sqrt{(y-y_k/d_l)^2+(z-z_k/d_l)^2}}}{\sqrt{1-2y\Phi_l-2z\Theta_l+y^2+z^2}}\cdot\cos\omega^{l,k}_{n_y,n_z}dydz\\&=\frac{1}{2}\int_0^{2\pi}\int_0^{+\infty}h_\vartheta(\rho)\cos(B_l\rho+\eta_{l,k})d\rho d\vartheta,\end{aligned} \tag{50}$$

$$\begin{aligned}[G]_{l,k} &= \sum_{n_y=-\frac{N_y-1}{2}}^{\frac{N_y-1}{2}}\sum_{n_z=-\frac{N_z-1}{2}}^{\frac{N_z-1}{2}}\frac{\lambda\sqrt{G_l}\sqrt{A}\cdot e^{-j\mathbf{k}_f\cdot\mathbf{d}^l_{n_y,n_z}}\cdot m_{n_y,n_z}\cdot e^{-\alpha|\mathbf{r}^k_{n_y,n_z}|}\cdot e^{-j\mathbf{k}_s\cdot\mathbf{r}^k_{n_y,n_z}}}{4\pi\sqrt{d_l^2-2d_l\Phi_l n_y d_y-2d_l\Theta_l n_z d_z+n_y^2d_y^2+n_z^2d_z^2}}\\&=\frac{\lambda\sqrt{G_l}\sqrt{A}e^{-j\mathbf{k}_f\cdot\mathbf{d}_l}}{4\pi d_l}\sum_{n_y=-\frac{N_y-1}{2}}^{\frac{N_y-1}{2}}\sum_{n_z=-\frac{N_z-1}{2}}^{\frac{N_z-1}{2}}\frac{e^{-\alpha\sqrt{(n_yd_y-y_k)^2+(n_zd_z-z_k)^2}-j\omega^{l,k}_{n_y,n_z}}\left[\sum_{l',k'}\frac{a_{l',k'}(1+\cos\omega^{l',k'}_{n_y,n_z})}{2}\right]}{\sqrt{1-2\Phi_l\frac{n_yd_y}{d_l}-2\Theta_l\frac{n_zd_z}{d_l}+(\frac{n_yd_y}{d_l})^2+(\frac{n_zd_z}{d_l})^2}}.\end{aligned} \tag{46}$$

$$[\mathbf{G}]_{l,k}=\frac{\lambda\sqrt{G_l}\sqrt{A}d_le^{-j\mathbf{k}_f\cdot\mathbf{d}_l}}{4\pi d_yd_z}\iint_{\mathcal{A}}\frac{e^{-\alpha d_l\sqrt{(y-y_k/d_l)^2+(z-z_k/d_l)^2}-j\omega_{l,k}}}{\sqrt{1-2y\Phi_l-2z\Theta_l+y^2+z^2}}\cdot\left[\sum_{l'=1}^{L}\sum_{k'=1}^{K}\frac{a_{l',k'}(1+\cos\omega^{l',k'}_{n_y,n_z})}{2}\right]dydz, \tag{48}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2022.3143221, IEEE Journal on Selected Areas in Communications

$$\begin{aligned}
&(\cos\omega_{n_y,n_z}^{l,k}+j\sin\omega_{n_y,n_z}^{l,k})\cdot\left[\sum_{l'=1}^{L}\sum_{k'=1}^{K}\frac{a_{l',k'}(1+\cos\omega_{n_y,n_z}^{l',k'})}{2}\right]\\
&=\frac{a_{l,k}}{4}+a_{l,k}\left(\frac{\cos\omega_{n_y,n_z}^{l,k}}{2}+\frac{\cos 2\omega_{n_y,n_z}^{l,k}}{4}\right)+\sum_{l'\neq l}\sum_{k'\neq k}a_{l',k'}\left[\frac{\cos\omega_{n_y,n_z}^{l,k}}{2}+\frac{\cos(\omega_{n_y,n_z}^{l,k}+\omega_{n_y,n_z}^{l',k'})}{4}\right.\\
&\left.+\frac{\cos(\omega_{n_y,n_z}^{l,k}-\omega_{n_y,n_z}^{l',k'})}{4}\right]+j\cdot\sum_{l',k'}a_{l',k'}\left[\frac{\sin\omega_{n_y,n_z}^{l,k}}{2}+\frac{\sin(\omega_{n_y,n_z}^{l,k}+\omega_{n_y,n_z}^{l',k'})}{4}+\frac{\sin(\omega_{n_y,n_z}^{l,k}-\omega_{n_y,n_z}^{l',k'})}{4}\right].
\end{aligned}\tag{49}$$

where $h_\vartheta(\rho)=\frac{\rho e^{-\alpha d_l\rho}}{\sqrt{(\rho\cos\vartheta+y_k/d_l-\Phi_l)^2+(\rho\sin\vartheta+z_k/d_l-\Theta_l)^2+\Psi_l^2}}$, $\Psi_l=\sin\theta_l\cos\varphi_l$, $B_l=d_l[|\mathbf{k}_s|-|\mathbf{k}_f|\cdot(\Phi_l\cos\vartheta+\Theta_l\sin\vartheta)]$, $\eta_{l,k}=|\mathbf{k}_f|(y_k\Phi_l+z_k\Theta_l)$. For a fixed ϑ, by utilizing integration by parts we have

$$\begin{aligned}
&\int_0^{+\infty}h_\vartheta(\rho)\cos(B_l\rho+\eta_{l,k})d\rho\\
&\overset{(a)}{=}-\frac{1}{B_l}\int_0^{+\infty}h_\vartheta(\rho)\sin(B_l\rho+\eta_{l,k})d\rho\\
&=\frac{1}{B_l^2}h'_\vartheta(0)\cos\eta_{l,k}-\frac{1}{B_l^2}\int_0^{+\infty}h''_\vartheta(\rho)\cos(B_l\rho+\eta_{l,k})d\rho\\
&=\frac{1}{B_l^2}\cdot\frac{\cos\eta_{l,k}}{[(y_k/d_l-\Phi_l)^2+(z_k/d_l-\Theta_l)^2+\Psi_l^2]^{\frac{3}{2}}}-\\
&\quad\frac{1}{B_l^3}\int_0^{+\infty}h''_\vartheta(\rho)d\sin(B_l\rho+\eta_{l,k}),
\end{aligned}\tag{51}$$

where in (a) we utilize $h_\vartheta(0)=h_\vartheta(+\infty)=0$. Since $B_l\geq d_l(|\mathbf{k}_s|-|\mathbf{k}_f|)$ has a magnitude of 10^3, the first item in (51) has a magnitude of 10^{-6} and the second item has a magnitude of 10^{-9}. The integral corresponding to $\frac{\cos\omega_{n_y,n_z}^{l,k}}{2}$ can then be neglected. This completes the proof.

Appendix C
Proof of Proposition 2

Without loss of generality, we assume that $I_{l,1}\geq I_{l,2}\geq\cdots\geq I_{l,K}$. Therefore, we have $\sum_{k=2}^K a_{l,k}^2 I_{l,k}^2\leq I_{l,1}^2\sum_{k=2}^K a_{l,k}^2\leq I_{l,1}^2\sum_{k=2}^K a_{l,k}\cdot(\sum_{k=1}^K a_{l,k})\leq I_{l,1}^2(a_{l,1}+\sum_{k=1}^K a_{l,k})\sum_{k=2}^K a_{l,k}$, which is equivalent to

$$\begin{aligned}
&\sum_{k=2}^K a_{l,k}^2 I_{l,k}^2\leq I_{l,1}^2(a_{l,1}+\sum_{k=1}^K a_{l,k})(-a_{l,1}+\sum_{k=1}^K a_{l,k})\\
&\Leftrightarrow\sum_{k=1}^K a_{l,k}^2 I_{l,k}^2\leq(\sum_{k=1}^K a_{l,k})^2 I_{l,1}^2,
\end{aligned}\tag{52}$$

where the equality holds when $a_{l,k}$ satisfies $\begin{cases}a_{l,k}\neq 0, & k=1,\\ a_{l,k}=0, & k\neq 1\end{cases}$. This completes the proof.

Appendix D
Proof of Proposition 3

We first prove that the solution given by (29) is the only interior maximum point of problem (19). Since the Hessian matrix of $\sum_{l=1}^L\log_2(1+I_l a_l^2)$ with respect to a_l is $\mathrm{diag}\{\frac{1-I_l a_l^2}{(1+I_l a_l^2)^2}\cdot\frac{2I_l}{\ln 2}\}$, the maximum point of problem (19) must satisfy $1-I_l a_l^2\leq 0$, $\forall l$. For each user l, the equation $\partial\mathcal{L}/\partial a_l=0$ has two solutions $a_l=\frac{1}{\beta\ln 2}\pm\sqrt{\frac{1}{(\beta\ln 2)^2}-\frac{1}{I_l}}$, and $1-I_l a_l^2$ can then be given by

$$1-I_l a_l^2=2\sqrt{\frac{I_l}{(\beta\ln 2)^2}-1}\left(\mp\sqrt{\frac{I_l}{(\beta\ln 2)^2}}-\sqrt{\frac{I_l}{(\beta\ln 2)^2}-1}\right).\tag{53}$$

Therefore, the maxima is achieved only when $a_l=\frac{1}{\beta\ln 2}+\sqrt{\frac{1}{(\beta\ln 2)^2}-\frac{1}{I_l}}$.

We now prove that when $\min_l I_l\geq 12(L-1)^2$, the solution given by (29) is the global maximum point of problem (19), i.e., the maxima can not be achieved on the boundary of the feasible region $\{\sum_{l=1}a_l=1, a_l\geq 0,\forall l\}$. We prove this by contradiction. Suppose there exists an optimal solution a_l such that $a_l=0$ for at least one user l. Without loss of generality, we assume that $I_1\leq I_2\leq\cdots\leq I_L$. Note that if $I_{l_1}\geq I_{l_2}$, then $a_{l_1}\geq a_{l_2}$, otherwise exchanging the values of a_{l_1} and a_{l_2} will lead to an increase of the capacity, which contradicts the optimality of a_l. Therefore, $a_1\leq a_2\leq\cdots\leq a_L$, and thus, $a_1=0$ and $a_L\geq\frac{1}{L-1}$. Define another feasible solution $b_l=\begin{cases}\frac{a_L}{2} & l=1,L,\\ a_l, & 2\leq l\leq L-1\end{cases}$. When $I_1\geq 12(L-1)^2$, we have

$$\begin{aligned}
&\sum_{l=1}^L\log_2(1+I_l b_l^2)-\sum_{l=1}^L\log_2(1+I_l a_l^2)\\
&=\log_2[(1+\frac{1}{4}I_1 a_L^2)(1+\frac{1}{4}I_L a_L^2)]-\log_2(1+I_L a_L^2)\\
&>\log_2(1+\frac{1}{4}I_L a_L^2+\frac{1}{16}I_L a_L^2\cdot I_1 a_L^2)-\log_2(1+I_L a_L^2)\\
&\geq\log_2(1+\frac{1}{4}I_L a_L^2+\frac{1}{16}I_L a_L^2\cdot 12)-\log_2(1+I_L a_L^2)=0,
\end{aligned}\tag{54}$$

indicating that the capacity corresponding to $\{b_l\}$ is larger than that of $\{a_l\}$, which contradicts the optimality of a_l. Therefore, the maxima can not be achieved on the boundary of the feasible region and the solution given by (29) is the global optimal solution to problem (19).

Appendix E
Proof of Proposition 5

According to (14) and (32), the data rate of user l is $R_l=\log_2\left(1+\frac{P_l}{\sigma^2}(\mathbf{G}_l\mathbf{V}_l)(\mathbf{G}_l\mathbf{V}_l)^H\right)$, where $\mathbf{G}_l$ is the l-th row of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2022.3143221, IEEE Journal on Selected Areas in Communications

$\mathbf{G}$. Based on (32), $(\mathbf{G}_l\mathbf{V}_l)(\mathbf{G}_l\mathbf{V}_l)^H$ can be rewritten as

$$(\mathbf{G}_l\mathbf{V}_l)(\mathbf{G}_l\mathbf{V}_l)^H = \mathbf{G}_l\mathbf{V}_l\mathbf{V}_l^H\mathbf{G}_l^H = \mathbf{G}_l\frac{\mathbf{W}_l}{|\mathbf{W}_l|}\frac{\mathbf{W}_l^H}{|\mathbf{W}_l|}\mathbf{G}_l^H. \tag{55}$$

Since $\mathbf{W} = \mathbf{G}^H(\mathbf{G}\mathbf{G}^H)^{-1}$, we have $\mathbf{G}_l\mathbf{W}_l\mathbf{W}_l^H\mathbf{G}_l^H = [\mathbf{G}\mathbf{G}^H(\mathbf{G}\mathbf{G}^H)^{-1}]_{l,l}[\mathbf{G}\mathbf{G}^H(\mathbf{G}\mathbf{G}^H)^{-1}]_{l,l}^H = 1$. Therefore, $(\mathbf{G}_l\mathbf{V}_l)(\mathbf{G}_l\mathbf{V}_l)^H$ can be further expressed as

$$\begin{aligned}(\mathbf{G}_l\mathbf{V}_l)(\mathbf{G}_l\mathbf{V}_l)^H &= \frac{1}{|\mathbf{W}_l|^2} = \frac{1}{[\mathbf{W}^H\mathbf{W}]_{l,l}} \\ &= \frac{1}{[(\mathbf{G}\mathbf{G}^H)^{-1}\mathbf{G}\mathbf{G}^H(\mathbf{G}\mathbf{G}^H)^{-1}]_{l,l}} = \frac{1}{[(\mathbf{G}\mathbf{G}^H)^{-1}]_{l,l}}.\end{aligned} \tag{56}$$

This completes the proof.

Appendix F
Proof of Lemma 3

Based on the expression of $I_{l,k}$ given in (22), for an extremely large-scale RHS satisfying $N_y \to \infty$ and $N_z \to \infty$, $I_{l,k}$ can be rewritten as $I_{l,k} = \frac{1}{d_l}\iint_{\mathbb{R}^2}\frac{e^{-\alpha\sqrt{u^2+v^2}}dudv}{\sqrt{(u+y_k-d_l\Phi_l)^2+(v+z_k-d_l\Theta_l)^2+(d_l\Psi_l)^2}}$, where $\Psi_l = \sin\theta_l\cos\varphi_l$. Utilizing the polar coordinates, we define $y_k - d_l\Phi_l = t_{l,k}\cos\delta_{l,k}$ and $z_k - d_l\Theta_l = t_{l,k}\sin\delta_{l,k}$. We then have $I_{l,k} = \frac{1}{d_l}\int_0^{2\pi}\int_0^{+\infty}\frac{\rho e^{-\alpha\rho}}{\sqrt{\rho^2+t_{l,k}^2+(d_l\Psi_l)^2+2\rho t_{l,k}\cos(\delta_{l,k}-\vartheta)}}d\vartheta d\rho$. Due to the periodicity of the cosine function, the value of $\delta_{l,k}$ does not influence $I_{l,k}$, and thus, we assume that $\delta_{l,k} = 0$. $I_{l,k}$ can then be simplified as $I_{l,k} = \frac{1}{d_l}\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty}\frac{e^{-\alpha\sqrt{u^2+v^2}}}{\sqrt{(u+t_{l,k})^2+v^2+(d_l\Psi_l)^2}}du\right)dv$. The derivative of $I_v(t_{l,k}) = \int_{-\infty}^{+\infty}\frac{e^{-\alpha\sqrt{u^2+v^2}}}{\sqrt{(u+t_{l,k})^2+v^2+(d_l\Psi_l)^2}}du$ can then be given by

$$\begin{aligned}I_v'(t_{l,k}) &= \int_{-\infty}^{+\infty}\frac{d}{dt_{l,k}}\left[\frac{e^{-\alpha\sqrt{u^2+v^2}}}{\sqrt{(u+t_{l,k})^2+v^2+(d_l\Psi_l)^2}}\right]du \\ &= \int_{-\infty}^{+\infty}\frac{-(u+t_{l,k})e^{-\alpha\sqrt{u^2+v^2}}}{[\sqrt{(u+t_{l,k})^2+v^2+(d_l\Psi_l)^2}]^{\frac{3}{2}}}du \\ &= \int_{-\infty}^{+\infty}\frac{-ue^{-\alpha\sqrt{(u-t_{l,k})^2+v^2}}}{[\sqrt{u^2+v^2+(d_l\Psi_l)^2}]^{\frac{3}{2}}}du \\ &= \int_{0}^{+\infty}\frac{u(e^{-\alpha\sqrt{(u+t_{l,k})^2+v^2}}-e^{-\alpha\sqrt{(u-t_{l,k})^2+v^2}})}{[\sqrt{u^2+v^2+(d_l\Psi_l)^2}]^{\frac{3}{2}}}du.\end{aligned} \tag{57}$$

Since $(u+t_{l,k})^2 \geq (u-t_{l,k})^2$ when $u \geq 0$ and $t_{l,k} \geq 0$, $I_v'(t_{l,k}) \leq 0$, and thus, $I_{l,k}$ decreases as $t_{l,k}$ grows. Therefore, $I_{l,k}$ decreases as $d_{l,k} = \sqrt{t_{l,k}^2+(d_l\Psi_l)^2}$ grows. This completes the proof.

Appendix G
Proof of Proposition 6

When the feed closest to each user is different, without loss generality, it can be assumed that $k_l^* = l, \forall l$. Bases on Proposition 3, we have

$$\begin{cases} a_{l,k} \neq 0, & k = l, \\ a_{l,k} = 0, & k \neq l. \end{cases} \Rightarrow \begin{cases} [\mathbf{G}]_{l,k} \neq 0, & k = l, \\ [\mathbf{G}]_{l,k} = 0, & k \neq l. \end{cases} \tag{58}$$

Therefore, $\mathbf{G}$ is a diagonal matrix, and thus, $\mathbf{G}\mathbf{G}^H$ is also a diagonal matrix satisfying $[(\mathbf{G}\mathbf{G}^H)^{-1}]_{l,l} = [(\mathbf{G}\mathbf{G}^H)]_{l,l}^{-1}$. According to (34), the sum rate with ZF precoding of the HDMA wireless communication system can be expressed as

$$\begin{aligned}R &= \sum_{l=1}^{L}R_l = \sum_{l=1}^{L}\log_2\left(1+\frac{P_l}{\sigma^2[(\mathbf{G}\mathbf{G}^H)^{-1}]_{l,l}}\right) \\ &= \sum_{l=1}^{L}\log_2\left(1+\frac{P_l}{\sigma^2}[(\mathbf{G}\mathbf{G}^H)]_{l,l}\right) = C.\end{aligned} \tag{59}$$

Since $\boldsymbol{m}$ given in (30) is the optimal holographic pattern maximizing the capacity, the sum rate with ZF precoding can achieve the asymptotic capacity.

Appendix H
Proof of Lemma 4

Based on the expression of I given in (41), I can be rewritten as (60), where the integral domain $\mathcal{A}' = \left\{(y,z)\middle|0 \leq y \leq \frac{N_yd_y}{2d_l}, 0 \leq z \leq \frac{N_zd_z}{2d_l}\right\}$. Define $f(x) = (1+x)^{-\frac{1}{2}} + (1-x)^{-\frac{1}{2}}$, where $|x| = |\frac{2y\Phi\pm2z\Theta}{1+y^2+z^2}| \leq \frac{2\sqrt{y^2+z^2}\sqrt{\Phi^2+\Theta^2}}{1+y^2+z^2}$. Since $f'(x) = \frac{1}{2}[(1-x)^{-\frac{3}{2}} - (1+x)^{-\frac{3}{2}}]$, $f'(x) < 0$ when $x < 0$, and $f'(x) > 0$ when $x > 0$, implying that $f(x)$ decreases monotonously when $x < 0$ and increases monotonously when $x > 0$. Therefore, the upper bound and lower bound of $f(x)$ can be given by $2 \leq f(x) \leq f(\pm\frac{2\sqrt{y^2+z^2}\sqrt{\Phi^2+\Theta^2}}{1+y^2+z^2})$. The upper bound and lower bound of I can then be obtained. This completes the proof.

Appendix I
Proof of Proposition 7 and Proposition 8

For an extremely large-scale RHS, the first item in (42) can be upper bounded by

$$\begin{aligned}&\frac{1}{2}\iint_{\mathbb{R}^2}\frac{e^{-\alpha d\sqrt{y^2+z^2}}}{\sqrt{1-2\sqrt{(y^2+z^2)(\Phi^2+\Theta^2)}+y^2+z^2}}dydz \\ &= \pi\int_0^{+\infty}\frac{\rho e^{-\alpha d\rho}}{\sqrt{1-2\Omega\rho+\rho^2}}d\rho = \pi\int_0^{+\infty}\frac{\rho e^{-\alpha d\rho}}{\sqrt{(\rho-\Omega)^2+\Psi^2}}d\rho \\ &= \pi\cdot e^{-\alpha d\Omega}\int_{-\Omega}^{+\infty}\frac{(\rho+\Omega)e^{-\alpha d\rho}}{\sqrt{\rho^2+\Psi^2}}d\rho \\ &\leq \pi\cdot e^{-\alpha d\Omega}\left(\int_{-\Omega}^{0}\frac{(\rho+\Omega)e^{-\alpha d\rho}}{\Psi}d\rho + \int_0^{+\infty}\frac{\rho e^{-\alpha d\rho}}{\sqrt{\rho^2+\Psi^2}}d\rho\right. \\ &\qquad\left. +\Omega\int_0^{+\infty}\frac{e^{-\alpha d\rho}}{\sqrt{\rho^2+\Psi^2}}d\rho\right) \\ &= \frac{\pi[1-(\alpha d\Omega+1)e^{-\alpha d\Omega}]}{\alpha^2d^2\Psi} + \pi e^{-\alpha d\Omega}(I_1+\Omega I_2),\end{aligned} \tag{61}$$

where $\Omega = \sqrt{\Phi^2+\Theta^2}$, $\Psi = \sin\theta\cos\varphi$. I_1 and I_2 can be expressed as

$$\begin{aligned}I_1 &= \int_0^{+\infty}\frac{\rho e^{-\alpha d\rho}}{\sqrt{\rho^2+\Psi^2}}d\rho \\ &= \frac{4-\pi\alpha d\Psi\mathrm{Y}_1(\alpha d\Psi)-\pi\alpha d\Psi\mathrm{H}_{-1}(\alpha d\Psi)}{2\alpha d},\end{aligned} \tag{62}$$

$$I = \iint_{\mathcal{A}'} \frac{e^{-\alpha d\sqrt{y^2+z^2}}}{\sqrt{1+y^2+z^2}} \left[\left(1 - \frac{2y\Phi + 2z\Theta}{1+y^2+z^2}\right)^{-\frac{1}{2}} + \left(1 + \frac{2y\Phi + 2z\Theta}{1+y^2+z^2}\right)^{-\frac{1}{2}} + \left(1 - \frac{2y\Phi - 2z\Theta}{1+y^2+z^2}\right)^{-\frac{1}{2}} + \left(1 + \frac{2y\Phi - 2z\Theta}{1+y^2+z^2}\right)^{-\frac{1}{2}} \right], \tag{60}$$

$$\begin{aligned}[\mathbf{G}]_{l,k} &= \frac{\lambda\sqrt{G_l}\sqrt{A}d_l e^{-j\mathbf{k}_f\cdot\mathbf{d}_l}}{4\pi d_y d_z} \iint_{\mathcal{A}} \frac{e^{-j\mathbf{k}_f\cdot\mathbf{r}_{n_y,n_z}}}{\sqrt{1-2y\Phi_l-2z\Theta_l+y^2+z^2}} \cdot \frac{e^{j\phi^k_{n_y,n_z}}}{\sqrt{N}} dydz \\ &= \frac{\lambda\sqrt{G_l}\sqrt{A}d_l e^{-j\mathbf{k}_f\cdot\mathbf{d}_l}}{4\pi d_y d_z\sqrt{N}} \iint_{\mathcal{A}} \frac{e^{-j\mathbf{k}_f\cdot\mathbf{r}_{n_y,n_z}}}{\sqrt{1-2y\Phi_l-2z\Theta_l+y^2+z^2}} \cdot e^{j\phi^k_{n_y,n_z}} dydz.\end{aligned} \tag{67}$$

$$I_2 = \int_0^{+\infty} \frac{e^{-\alpha d\rho}}{\sqrt{\rho^2+\Psi^2}} d\rho = \frac{\pi}{2}[\mathrm{H}_0(\alpha d\Psi) - \mathrm{Y}_0(\alpha d\Psi)], \tag{63}$$

where $\mathrm{Y}_\nu(\cdot)$ is the Bessel function of the second kind, $\mathrm{H}_\nu(\cdot)$ is the Struve function. The second item in (42) can be upper bounded by

$$\begin{aligned}&\frac{1}{2}\iint_{\mathbb{R}^2} \frac{e^{-\alpha d\sqrt{y^2+z^2}}}{\sqrt{1+2\sqrt{(y^2+z^2)(\Phi^2+\Theta^2)}+y^2+z^2}} dydz \\ &\leq \frac{1}{2}\iint_{\mathbb{R}^2} \frac{e^{-\alpha d\sqrt{y^2+z^2}}}{\sqrt{1+y^2+z^2}} dydz = \pi\int_0^{+\infty} \frac{\rho e^{-\alpha d\rho}}{\sqrt{1+\rho^2}} d\rho \\ &= \pi\frac{4-\pi\alpha d\mathrm{Y}_1(\alpha d) - \pi\alpha d\mathrm{H}_{-1}(\alpha d)}{2\alpha d}.\end{aligned} \tag{64}$$

Based on (61) and (64), the upper bound of the data rate in the single-user case given by (40) can be obtained as

$$\begin{aligned}R_s \leq \log_2\Bigg(1 + \frac{P\lambda^2 GAd^2}{256\sigma^2 d_y^2 d_z^2}\Big[e^{-\alpha d\Omega}(I_1 + \Omega I_2) + \\ \frac{1-(\alpha d\Omega+1)e^{-\alpha d\Omega}}{\alpha^2 d^2\Psi} + \frac{4-\pi\alpha d\mathrm{Y}_1(\alpha d) - \pi\alpha d\mathrm{H}_{-1}(\alpha d)}{2\alpha d}\Big]^2\Bigg),\end{aligned} \tag{65}$$

where I_1 and I_2 are given in (62) and (63), respectively. Note that the lower bound of I given in (42) can be directly derived from (64). Therefore, the lower bound of the data rate can be obtained as

$$\begin{aligned}R_s \geq \log_2\Bigg\{1 + \frac{P\lambda^2 GAd^2}{256\sigma^2 d_y^2 d_z^2}\cdot \\ \left[\frac{4-\pi\alpha d\mathrm{Y}_1(\alpha d) - \pi\alpha d\mathrm{H}_{-1}(\alpha d)}{\alpha d}\right]^2\Bigg\}.\end{aligned} \tag{66}$$

APPENDIX J
PROOF OF PROPOSITION 9

In the HDMA system, since when the physical dimension of the RHS along the y-axis and z axis are fixed, i.e., $N_y d_y$ and $N_z d_z$ are constants, the integral domain $\mathcal{A}$ in (14) is fixed and the integrand is independent of d_y and d_z. Therefore, the integral in (17) is independent of d_y and d_z, and thus, $[\mathbf{G}]_{l,k} \propto (d_y d_z)^{-1}$, indicating that $[\mathbf{G}\mathbf{G}^H]_{l,l} = \sum_{k=1}^K |[\mathbf{G}]_{l,k}|^2 \propto (d_y d_z)^{-1}$. Since the capacity C is $\sum_{l=1}^L \log_2\left(1 + \frac{P_l}{\sigma^2}[\mathbf{G}\mathbf{G}^H]_{l,l}\right)$ and $\frac{P_l}{\sigma^2}[\mathbf{G}\mathbf{G}^H]_{l,l} \gg 1$, C decreases logarithmically as the inverse square of the element spacing d_y and d_z grows, i.e., $C \propto \log_2(d_y d_z)^{-1}$. Moreover, since element number $N = N_y N_z \propto (d_y d_z)^{-1}$, the capacity also increases logarithmically with the element number N, i.e., $C \propto \log_2 N$.

In the traditional SDMA system, similar to the proof given in Appendix 1 (i.e., Proposition 1), the equivalent channel $[\mathbf{G}]_{l,k}$ between the RF chain k and user l can be expressed as (67) [27], where $\mathcal{A} = \left\{(y,z)\middle| |y| \leq \frac{N_y d_y}{2d_l}, |z| \leq \frac{N_z d_z}{2d_l}\right\}$ is the integral domain, $\{\frac{e^{j\phi^k_{n_y,n_z}}}{\sqrt{N}}\}$ is the precoding matrix in SDMA [10]. When the physical dimension of the RHS along the y-axis and z-axis are fixed, i.e., $N_y d_y$ and $N_z d_z$ are constants, the integral domain $\mathcal{A}$ is fixed and the integrand is independent of d_y and d_z. Therefore, the integral in (67) is independent of N_y and N_z, and thus, $[\mathbf{G}]_{l,k} \propto (N)^{\frac{1}{2}}$, indicating that $[\mathbf{G}\mathbf{G}^H]_{l,l} = \sum_{k=1}^K |[\mathbf{G}]_{l,k}|^2 \propto N$. Since the capacity C is $\sum_{l=1}^L \log_2\left(1 + \frac{P_l}{\sigma^2}[\mathbf{G}\mathbf{G}^H]_{l,l}\right)$ and $\frac{P_l}{\sigma^2}[\mathbf{G}\mathbf{G}^H]_{l,l} \gg 1$, C increases logarithmically with the element number N, i.e., $C \propto \log_2 N$. This completes the proof.

REFERENCES

[1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186-195, Feb. 2014.

[2] G. Xu and S. Li, "Throughput multiplication of wireless LANs for multimedia services: SDMA protocol design," *IEEE Global Commun. Conf. (GLOBECOM)*, 1994, pp. 1326-1332, vol.3.

[3] D. Gesbert, M. Shafi, Da-shan Shiu, P. J. Smith, and A. Naguib, "From theory to practice: an overview of MIMO space-time coded wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 281-302, Apr. 2003.

[4] J. Mietzner, R. Schober, L. Lampe, W. H. Gerstacker, and P. A. Hoeher, "Multiple-antenna techniques for wireless communications - a comprehensive literature survey," *IEEE Commun. Surveys Tuts.*, vol. 11, no. 2, pp. 87-105, Second Quarter, 2009.

[5] Z. Ding and H. V. Poor, "Design of massive-MIMO-NOMA with limited feedback," *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 629-633, May 2016.

[6] D. Wang, Y. Zhang, H. Wei, X. You, X. Gao, and J. Wang, "An overview of transmission theory and techniques of large-scale antenna systems for 5G wireless communications," *Sci. China Inf. Sci.* vol. 59, no. 8, pp. 1-18, Aug. 2016.

[7] T. Wang, S. Wang, and Z. Zhou, "Machine learning for 5G and beyond from model-based to data-driven mobile wireless networks," *China Commun.*, vol. 16, no. 1, pp. 165-175, Jan. 2019.

[8] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40-60, Jan. 2013.

[9] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 172-179, Feb. 2013.

[10] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501-513, Apr. 2016.

[11] N. Zheludev, "The road ahead for metamaterials," *Science*, vol. 328, no. 5978, pp. 582-583, Apr. 2010.

[12] N. Zheludev and Y. Kivshar, "From metamaterials to metadevices," *Nature Mater.*, vol. 11, no. 11, pp. 917-927, Oct. 2012.

[13] R. Liu, T. Cui, D. Huang, B. Zhao, and D. R. Smith, "Description and explanation of electromagnetic behaviors in artificial metamaterials based on effective medium theory", *Physical Review E*, vol. 76, no. 2, Aug. 2007.

[14] R. B. Hwang, "Binary meta-hologram for a reconfigurable holographic metamaterial antenna," *Sci. Rep.*, vol. 10, no. 1, pp. 1-10, May. 2020.

[15] T. Sleasman, M. F. Imani, W. Xu, J. Hunt, T. Driscoll, M. S. Reynolds, and D. R. Smith, "Waveguide-fed tunable metamaterial element for dynamic apertures," *IEEE Antennas Wireless Propag. Lett.*, vol. 15, pp. 606-609, July 2016.

[16] R. Deng, B. Di, H. Zhang, Y. Tan, and L. Song, "Reconfigurable holographic surface: Holographic beamforming for metasurface-aided wireless communications," *IEEE Trans. Vehicular Tech.*, vol. 70, no. 6, pp. 6255-6259, June 2021.

[17] Y. Liu *et al.*, "Reconfigurable intelligent surfaces: Principles and opportunities," *IEEE Commun. Surveys Tuts.*, Early access, doi: 10.1109/COMST.2021.3077737.

[18] B. Di, "Reconfigurable Holographic metasurface aided wideband OFDM communications against beam squint," *IEEE Trans. Vehicular Tech.*, vol. 70, no. 5, pp. 5099-5103, May 2021.

[19] B. Di, H. Zhang, L. Song, Y. Li, Z. Han, and H. V. Poor, "Hybrid beamforming for reconfigurable intelligent surface based multi-user communications: Achievable rates with limited discrete phase shifts," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1809-1822, Aug. 2020.

[20] M. Nemati, B. Maham, S. R. Pokhrel, and J. Choi, "Modeling RIS empowered outdoor-to-indoor communication in mmWave cellular networks," 2021, *arXiv:2101.00736*.

[21] D. R. Smith, O. Yurduseven, L. P. Mancera, P. Bowen, and N. B. Kundtz, "Analysis of a waveguide-fed metasurface antenna," *Physical Review Applied*, vol. 8, no. 5, pp. 054048, Nov. 2017.

[22] "Holographic beamforming and phased arrays," Pivotal Commware, Kirkland, WA, 2019.

[23] Kymeta, "Metamaterial-surface flat-panel antenna technology," available at: *https://www.kymetacorp.com/wp-content/uploads/2019/06/Metamaterial-Surface-Antenna-Technology.pdf*, Jun., 2019.

[24] M. Johnson, S. Brunton, N. Kundtz, and N. Kutz, "Extremum-seeking control of the beam pattern of a reconfigurable holographic metamaterial antenna," *J. Opt. Soc. Amer. A*, vol. 33, no. 1, pp. 59-68, Jan. 2016.

[25] Y. Ma and J. Wang, "Theoretical modeling and analysis of circularly polarized annular leaky-wave antenna based on travelling-wave structure," *IEEE Access*, vol. 9, pp. 29392-29400, 2021.

[26] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160-171, Feb. 2013.

[27] H. Liu and Y. Zeng, "Communicating with extremely large-scale array/surface: Unified modelling and performance analysis," 2021, *arXiv:2104.13162*.

[28] B. Friedlander, "Localization of signals in the near-field of an antenna array," *IEEE Trans. Signal Process.*, vol. 67, no. 15, pp. 3885-3893, Aug. 2019.

[29] J. Yuan, H. Q. Ngo, and M. Matthaiou, "Towards large intelligent surface (LIS)-based communications," *IEEE Trans. Commun.*, vol. 68, no. 10, pp. 6568-6582, Oct. 2020.

[30] S. Hu, F. Rusek and O. Edfors, "Beyond massive MIMO: The potential of data transmission with large intelligent surfaces," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2746-2758, May, 2018.

[31] Q. Wang, I. Balasingham, M. Zhang and X. Huang, "Improving RSS-Based ranging in LOS-NLOS scenario using GMMs," *IEEE Communications Lett.*, vol. 15, no. 10, pp. 1065-1067, Oct. 2011.

[32] L. Tran, M. Juntti, M. Bengtsson, and B. Ottersten, "Weighted sum rate maximization for MIMO broadcast channels using dirty paper coding and zero-forcing methods," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2362-2373, June 2013.

[33] H. Zhang, B. Di, Z. Han, H. V. Poor, and L. Song, "Reconfigurable intelligent surface assisted multi-user communications: How many reflective elements do we need?," *IEEE Wireless Commun. Lett.*, vol. 10, no. 5, pp. 1098-1102, May 2021.

[34] Andrews and Larry. C, *Special functions of mathematics for engineers*. Bellingham, WA: Spie Press, 1998.

[35] F. Xu, K. Wu, and X. Zhang, "Periodic leaky-wave antenna for millimeter wave applications based on substrate integrated waveguide," *IEEE Trans. Antennas Propag.*, vol. 58, no. 2, pp. 340-347, Feb. 2010.

[36] *Study on New Radio Access Technology: Radio Frequency (RF) and Coexistence Aspects Release 14*, document 3GPP TR 38.803, Sept. 2017.

Ruoqi Deng (S'19) received the B.S. degree in electronic engineering from Peking University, China, in 2019, where she is currently pursuing the Ph.D. degree with the Department of Electronics. Her current research interests include reconfigurable holographic surface, integrated aerial access, and satellite networks.

Boya Di (S'17-M'19) (S'17-M'19) obtained her PhD degree from the Department of Electronics, Peking University, China, in 2019. Prior to that, she received the B.S. degree in electronic engineering from Peking University in 2014. She was a postdoc researcher at Imperial College London and is now an assistant professor at Peking University. Her current research interests include reconfigurable intelligent surfaces, multi-agent systems, edge computing, vehicular networks, and aerial access networks. She received the best doctoral thesis award from China Education Society of Electronics in 2019. She is also the recipient of 2021 IEEE ComSoc Asia-Pacific Outstanding Paper Award. She serves as an associate editor for IEEE Transactions on Vehicular Technology since June 2020. She has also served as a workshop co-chair for IEEE WCNC 2020&2021.

Hongliang Zhang (S'15-M'19) received the B.S. and Ph.D. degrees at the School of Electrical Engineering and Computer Science at Peking University, in 2014 and 2019, respectively. He was a Postdoctoral Fellow in the Electrical and Computer Engineering Department at the University of Houston, Texas. Currently, he is a Postdoctoral Associate in the Department of Electrical and Computer Engineering at Princeton University, New Jersey. His current research interest includes reconfigurable intelligent surfaces, aerial access networks, optimization theory, and game theory. He received the best doctoral thesis award from Chinese Institute of Electronics in 2019. He is an exemplary reviewer for IEEE Transactions on Communications in 2020. He is also the recipient of 2021 IEEE Comsoc Heinrich Hertz Award for Best Communications Letters and 2021 IEEE ComSoc Asia-Pacific Outstanding Paper Award. He has served as a TPC Member for many IEEE conferences, such as Globecom, ICC, and WCNC. He is currently an Editor for IET Communications and Frontiers in Signal Processing. He also serves as a Guest Editor for IEEE IoT-J special issue on Internet of UAVs over Cellular Networks.

Lingyang Song (S'03-M'06-SM'12-F'19) received his Ph.D. from the University of York, United Kingdom, in 2007, where he received the K. M. Stott Prize for excellent research. He worked as a research fellow at the University of Oslo, Norway, until rejoining Philips Research UK in March 2008. In May 2009, he joined the Department of Electronics, School of Electronics Engineering and Computer Science, Peking University, and is now a Boya Distinguished Professor. His main research interests include wireless communication and networks, signal processing, and machine learning. He was the recipient of the IEEE Leonard G. Abraham Prize in 2016 and the IEEE Asia Pacific (AP) Young Researcher Award in 2012. He has been an IEEE Distinguished Lecturer since 2015.