

Electromagnetic Hybrid Beamforming for Holographic MIMO Communications

Ran Ji, *Graduate Student Member, IEEE*, Chongwen Huang[✉], *Member, IEEE*,
Xiaoming Chen[✉], *Senior Member, IEEE*, Wei E. I. Sha[✉], *Senior Member, IEEE*,
Linglong Dai[✉], *Fellow, IEEE*, Jiguang He[✉], *Senior Member, IEEE*,
Zhaoyang Zhang[✉], *Senior Member, IEEE*, Chau Yuen[✉], *Fellow, IEEE*,
and Mérouane Debbah[✉], *Fellow, IEEE*

Abstract—It is well known that there is inherent radiation pattern distortion for the commercial base station antenna array, which usually needs three antenna sectors to cover the whole space. To eliminate pattern distortion and further enhance beamforming performance, we propose an electromagnetic hybrid beamforming (EHB) scheme based on a three-dimensional (3D) superdirective holographic antenna array. Specifically, EHB consists of antenna excitation current vectors (analog beamforming)

and digital precoding matrices, where the implementation of analog beamforming involves the real-time adjustment of the radiation pattern to adapt it to the dynamic wireless environment. Meanwhile, the digital beamforming is optimized based on the channel characteristics of analog beamforming to further improve the achievable rate of communication systems. An electromagnetic channel model incorporating array radiation patterns and the mutual coupling effect is also developed to evaluate the benefits of our proposed scheme. Simulation results demonstrate that our proposed EHB scheme with a 3D holographic array achieves a relatively flat superdirective beamforming gain and allows for programmable focusing directions throughout the entire spatial domain. Furthermore, they also verify that the proposed scheme achieves a sum rate gain of over 150% compared to traditional beamforming algorithms.

Index Terms—Mutual coupling, superdirectivity, holographic antenna array, electromagnetic hybrid beamforming.

I. INTRODUCTION

5G COMMUNICATIONS have greatly benefited from the massive multiple-input-multiple-output (MIMO) and beamforming technologies, which have been proven to be effective in enhancing communication rates and reliability, leading to significant advancements compared with 4G communications. However, the radiation pattern of massive MIMO antenna arrays in commercial base stations has a narrow angular spread of approximately ± 60 degrees [1, Fig. 4.26], which usually have the inherent radiation pattern distortion (i.e., radiation pattern that is non-uniform in the angular domain), especially for the directions of edge users, where the quality of service is relatively poor. In other words, a typical commercial antenna array usually exhibits a radiation gain loss of more than 4 dB (up to 20 dB especially for fixed broadside beamforming) in the edge directions, i.e., 60-degree direction, compared with the 0-degree (normal direction) [1, Fig. 6.7]. It is natural to think if it is possible to achieve a consistent and high-gain radiation pattern for users in all directions by leveraging a programmable radiation pattern and beamforming approaches.

A. Prior Works

Hybrid beamforming in millimeter wave or terahertz communication systems can partially address this issue, but its primary objectives remain to reduce hardware cost and

Manuscript received 13 February 2024; revised 31 May 2024 and 17 July 2024; accepted 23 July 2024. Date of publication 6 August 2024; date of current version 13 November 2024. This work was supported in part by China National Key Research and Development Program under Grant 2021YFA1000500 and Grant 2023YFB2904804; in part by the National Natural Science Foundation of China under Grant 62331023, Grant 62101492, Grant 62394292, and Grant U20A20158; in part by Zhejiang Provincial Natural Science Foundation of China under Grant LR22F010002; in part by Zhejiang Provincial Science and Technology Plan Project under Grant 2024C01033; in part by Zhejiang University Global Partnership Fund; in part by the Ministry of Education (MOE), Singapore, through its MOE Tier 2, under Award MOE-T2EP50220-0019; and in part by the Agency for Science, Technology and Research (ASTAR), Singapore, under Grant M22L1b0110. The associate editor coordinating the review of this article and approving it for publication was J. A. Zhang. (*Corresponding author: Chongwen Huang.*)

Ran Ji and Chongwen Huang are with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China, also with the State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an 710071, China, and also with Zhejiang Provincial Key Laboratory of Information Processing, Communications and Networks (IPCAN), Hangzhou 310027, China (e-mail: ranji@zju.edu.cn; chongwenhuang@zju.edu.cn).

Xiaoming Chen is with the School of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: xiaoming.chen@mail.xjtu.edu.cn).

Wei E. I. Sha is with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: weisha@zju.edu.cn).

Linglong Dai is with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: daill@tsinghua.edu.cn).

Jiguang He is with the Technology Innovation Institute, Abu Dhabi, United Arab Emirates (e-mail: jiguang.he@tii.ae).

Zhaoyang Zhang is with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China, and also with Zhejiang Provincial Key Laboratory of Information Processing, Communications and Networks (IPCAN), Hangzhou 310027, China (e-mail: zhzy@zju.edu.cn).

Chau Yuen is with the School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore 639798 (e-mail: chau.yuen@ntu.edu.sg).

Mérouane Debbah is with the KU 6G Research Center, Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates (e-mail: merouane.debbah@ku.ac.ae).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TWC.2024.3435844>.

Digital Object Identifier 10.1109/TWC.2024.3435844

optimize both the digital and analog beamforming matrices to approach the performance of digital solutions. Specifically, Huang et al. [2] proposed a hybrid beamforming architecture that combines low-cost analog radio-frequency (RF) chains with digital baseband precoding. Subsequently, Orthogonal Matching Pursuit (OMP) was employed for sparse channel reconstruction [3]. Moreover, corresponding channel estimation and beamforming algorithm designs based on a hierarchical multi-resolution codebook were also provided. Yu et al. [4] evaluated the performance of fully-connected and partially-connected hybrid beamforming schemes based on the principle of alternating minimization. By contrast, a sub-connected architecture was proposed in [5], where the successive interference cancellation method is employed to derive closed-form digital and analog precoders. Li and Alkhateeb [6] introduced a deep learning approach to hybrid beamforming design with channel encoder and precoder networks to perform compressive channel sensing and hybrid beam prediction, respectively. Moreover, a dynamic array-of-subarray architecture was further proposed, where the connections between analog precoders and RF chains can be dynamically adjusted to be adaptive to different channel conditions [7]. Although the analog part of hybrid beamforming can optimize the radiation pattern to some extent, it cannot completely eliminate the problem of pattern distortion.

To solve this problem, there have been some works showing that superdirectivity can be a potential approach to address the issue of distortion in edge radiation patterns. Specifically, by employing highly directive spatial beams, it is possible to achieve minor sidelobe energy leakage and improve the array gain. In superdirectivity antenna theory, excitation current vector and array geometry design have been investigated in earlier research [8], [9], [10], [11], [12]. It was demonstrated in [8] and [9] that the directivity of a linear array of M isotropic antennas can reach M^2 as the spacing between antennas approaches zero. Furthermore, the near-field characteristics of superdirectivity were investigated in [13] from a linear transformation perspective. Subsequent research in this field [10], [11], [12] mainly focused on the excitation coefficient design for superdirective beams under different application limitations. Although several limitations of superdirective antenna arrays have been identified, such as low radiation efficiency, high excitation precision requirements, and low bandwidth, the mutual coupling effect for antenna elements with intervals less than half the wavelength is ignored in the aforementioned works. In 2019, Marzetta [14] introduced the concept of superdirective arrays into communication systems. For further improved beamforming gain, subsequent works [15], [16], [17] proposed mutual coupling-based superdirective antenna arrays. Specifically, in the proposed analog beamforming technique, mutual coupling effects were utilized to generate superdirective spatial beams [15], [16], which is a desirable radiation pattern for further improvement of system performance. Through simulations and experimental measurements, M^2 directivity has been verified in the end-fire direction for linear arrays [16]. Although superdirectivity can address the pattern distortion issue for edge users, it may not provide coverage for other spatial regions.

Furthermore, superdirectivity requires densely placed antenna elements, which aligns with the recently proposed concept of holographic MIMO (HMIMO). HMIMO is an advanced technology building upon the concept of massive MIMO and involves integrating a massive (possibly infinite) number of antennas into a compact space. The pioneering works [18] and [19] provided an initial introduction to the emerging HMIMO wireless communication, in particular the available HMIMO hardware architectures, functionalities and characteristics. Such surfaces comprising dense electromagnetic (EM) excited elements are capable of recording and manipulating impinging fields with utmost flexibility and precision, in addition to reduced cost and power consumption, thereby shaping arbitrary-intended EM waves with high energy efficiency. Furthermore, a comprehensive overview of the latest advances in the HMIMO communications paradigm is presented in [20], with a special focus on their physical aspects, theoretical foundations, as well as the enabling technologies for HMIMO systems.

In particular, the degree-of-freedom (DOF) characteristics of HMIMO were investigated in [21] for far-field communication scenarios, revealing the optimized spatial sampling locations of electromagnetic fields on HMIMO surfaces. To enable near-field HMIMO communications, the additional near-field DOF performance was revealed in [22]. Moreover, [23], [24], and [25] further investigated near-field characteristics and performances of HMIMO systems. Additionally, the mutual coupling effect between antenna elements is prevalent in dense HMIMO arrays and the impact of mutual coupling on array radiation pattern and HMIMO performance was numerically studied in [26]. For the hardware design and implementation, an electromagnetic modulation scheme of HMIMO were proposed in [27], which is enabled by metamaterial antennas with tunable radiated amplitudes. Moreover, a holographic hybrid beamforming scheme was proposed in [28] and [29] based on the aforementioned HMIMO hardware.

B. Our Contributions

In this paper, we introduce an electromagnetic hybrid beamforming (EHB) scheme based on three-dimensional (3D) antenna array architecture into HMIMO communications to implement a programmable array radiation pattern in any desired direction with minor radiation pattern distortion. Then, an electromagnetic channel model and corresponding beamforming algorithms are investigated to improve the spectral efficiency. The contributions of the paper are listed as follows.

- We first propose an EHB scheme with a holographic antenna array. Specifically, the EHB scheme consists of antenna excitation current vectors (analog beamforming) and digital precoding matrices. The implementation of analog beamforming enables the real-time adjustment of the radiation pattern to achieve an approximately constant gain in different intended directions and the digital beamforming is optimized based on the channel characteristics of analog beamforming. Moreover, we also develop electromagnetic signal and channel models in EHB communication schemes, which provide explanations of the modulation effects of the EHB scheme

and the distribution of emitted energy from different perspectives.

- We investigate the hardware design of superdirectivity and 3D array geometry. Additionally, the optimal geometric arrangement of a 3D antenna array is also presented. Utilizing the 3D holographic antenna array structure and the customizable spatial radiation pattern, multiple superdirective beams in arbitrary directions can be generated by leveraging analog excitation vectors and taking the mutual coupling effects into account.
- Simulation results are finally provided to demonstrate that our proposed EHB scheme with a 3D holographic array achieves a relatively flat superdirective beamforming gain throughout the entire spatial domain. A sum rate gain of over 150% compared to traditional beamforming algorithms is also obtained by our proposed EHB.

C. Organization and Notation

1) *Organization*: The remainder of this paper is organized as follows. In Section II, we introduce the electromagnetic system model and the derivation of mutual coupling matrix. Then, an electromagnetic channel model is developed for further investigation. In Section III, we present the design principles of 3D antennas and the comprehensive EHB scheme, along with the corresponding algorithms. Numerical results are presented in Section IV to evaluate the effectiveness of our proposed EHB scheme with 3D holographic array structure. Conclusions are drawn in Section V.

2) *Notation*: Fonts a , \mathbf{a} and \mathbf{A} represent scalars, vectors and matrices, respectively. $\mathbf{A}^T, \mathbf{A}^H, \mathbf{A}^{-1}, \mathbf{A}^\dagger$ denote transpose, Hermitian, inverse and Moore-Penrose inverse operation respectively. $\|\mathbf{A}\|_F$ represents Frobenius norm of matrix \mathbf{A} . Operators $|\cdot|$ and $(\cdot)^*$ denote the modulus and conjugate operation respectively. Operators $\text{Tr}(\cdot)$ and $\text{diag}(\cdot)$ represent calculating the trace of a matrix and expanding a vector into a diagonal matrix, respectively.

II. ELECTROMAGNETIC SYSTEM MODEL

Different from the conventional hybrid beamforming scheme, we leverage the programmable electromagnetic characteristics of the holographic antenna array to solve the radiation pattern distortion. To this end, three challenges arise. *First*, it is non-trivial to accurately model the array radiation pattern of a densely placed holographic antenna array due to the mutual coupling [30], [31], [32]. *Second*, there is a need for further investigation into the impact of variations in the array radiation pattern on the performance of wireless communication systems. *Third*, although superdirective beams exhibit high directional gain, they can only be generated in end-fire directions for linear arrays [16], [17], which limits its practical application. The solutions to these three challenges will be discussed in the following two sections.

In this section, we present the electromagnetic transmit signal and channel models in HMIMO communications.

A. Electromagnetic Transmission Signal Model

For densely placed holographic arrays, the effect of mutual coupling between the antennas can cause significant variations

in the magnitude and phase characteristics of radiation patterns, resulting in deviations from the ideal analytical expressions. To accurately model the impact of these electromagnetic effects on signal transmission, we develop the following electromagnetic communication model.

Consider a HMIMO communication system depicted in Fig. 1, where N_T transmit antennas are densely arranged within a fixed array aperture. The adaptability of this hardware structure can be achieved with two components: digital beamforming \mathbf{W}_{BB} through RF chains and analog beamforming \mathbf{i} through antenna excitation currents. Firstly, we consider the conventional scenario where the antenna elements in the transmit MIMO array are assumed to be uncoupled. When the transmit signal \mathbf{x} is emitted through the holographic array transmitter, it naturally undergoes modulation by the pattern function of each antenna element, which describes the power distribution of its spatial radiation. Therefore, we can model the emitted signal as:

$$\mathbf{x}^{(e)} = \mathbf{i} \odot \mathbf{g}(\mathbf{r}, \boldsymbol{\theta}, \boldsymbol{\phi}) \odot \mathbf{x} = \mathbf{i} \odot \mathbf{g}(\mathbf{r}, \boldsymbol{\theta}, \boldsymbol{\phi}) \odot \mathbf{W}_{BB} \mathbf{t}, \quad (1)$$

where \odot represents element-wise product; $\mathbf{t} \in \mathbb{C}^{K \times 1}$, $\mathbf{W}_{BB} \in \mathbb{C}^{N_T \times K}$, $\mathbf{i} \in \mathbb{C}^{N_T \times 1}$ and $\mathbf{g}(\mathbf{r}, \boldsymbol{\theta}, \boldsymbol{\phi})$ represent the symbol vector for K users, the digital beamforming matrix, the antenna excitation vector and a $N_T \times 1$ vector function representing the radiation pattern of antenna elements, respectively. The last one can be formulated as

$$\mathbf{g}(\mathbf{r}, \boldsymbol{\theta}, \boldsymbol{\phi}) = [g(r_1, \theta_1, \phi_1), \dots, g(r_n, \theta_n, \phi_n), \dots, g(r_{N_T}, \theta_{N_T}, \phi_{N_T})]^T, \quad (2)$$

where r_n , θ_n , and ϕ_n represent the distance, elevation angle, and azimuth angle, respectively, of the measurement point relative to the n th transmit antenna. Note that in the far-field communication scenario, the parameter r_n can be disregarded, and θ_n and ϕ_n remain the same for all transmit antennas.

Secondly, we model the impact of the non-negligible mutual coupling effect on the transmission signal. After introducing the mutual coupling matrix [15], the distorted transmission signal due to the mutual coupling effect can be written as

$$\mathbf{x}^{(c)} = \mathbf{C} \mathbf{x}^{(e)} = \mathbf{C} \mathbf{i} \odot \mathbf{g}(\mathbf{r}, \boldsymbol{\theta}, \boldsymbol{\phi}) \odot \mathbf{W}_{BB} \mathbf{t}, \quad (3)$$

where c_{mn} represents the coherence of the n th element on the m th element.

B. Radiation Pattern Model

In this subsection, we investigate the mutual coupling matrix introduced in Section II. A in detail to model the actual radiation pattern of the densely placed holographic antenna array. Although some research has attempted to derive analytical expressions for mutual coupling between closely spaced antennas [33], [34], [35], [36], practical results often exhibit significant deviations from these analytical expressions when the idealized conditions for theoretical derivations are not met. Therefore, in order to accurately evaluate the effects of mutual coupling on practical radiation patterns, we obtain the coupling matrix based on full-wave electromagnetic field simulation results in both coupled and uncoupled scenarios, as well as spherical wave expansions. This approach decomposes the

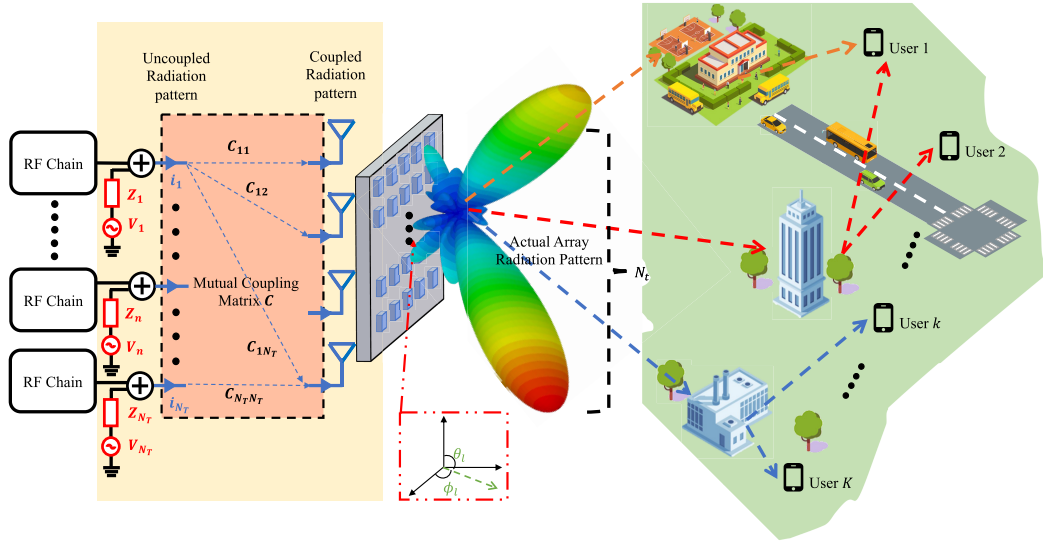


Fig. 1. Electromagnetic hybrid beamforming architecture.

electromagnetic field into a series of spherical wave coefficients using a set of orthogonal spherical wave bases.

The actual radiation pattern of the m th element can be written as [16], [36]:

$$g_{m,r}(r, \theta, \phi) = \sum_{n=1}^{N_T} c_{mn} i_n g_{n,i}(r, \theta, \phi), \quad (4)$$

where the first subscript represents the antenna index, while the second subscript indicates either the realized radiation pattern (r) or the ideal radiation pattern (i).

The coupling matrix can be derived from spherical wave expansion, with the correlation matrix coefficients obtained by full-wave simulation results. With this expansion, the electric field $\mathbf{E}(r, \theta, \phi)$ radiated by one antenna element can be expanded as [37]:

$$\begin{aligned} e(r, \theta, \phi) &= \kappa \sqrt{\eta} \sum_{s=1}^2 \sum_{n=1}^N \sum_{m=-n}^n Q_{s,m,n} \mathbf{F}_{s,m,n}(r, \theta, \phi) \\ &= \kappa \sqrt{\eta} \sum_{s,m,n} Q_{s,m,n} \mathbf{F}_{s,m,n}(r, \theta, \phi), \end{aligned} \quad (5)$$

where κ is the wavenumber and η is the propagation medium intrinsic impedance. $Q_{s,m,n}$ is the spherical wave expansion coefficient and s, m, n denotes electrical mode (TE/TM), degree of the wave and order of the wave, respectively. N is the maximum order of available spherical wave functions,

and is determined by the array aperture size. According to Chu-Harrington Limitation [38], [39], the available order of spherical wave expansion is determined by the electrical size of the transmitter as $N \approx \kappa a$, where a is the minimum radius of a spherical surface which can enclose the antenna. To be specific, applying an order larger than κa will result in an exponentially increasing Q factor, bringing a narrow-banded antenna. Moreover, $\mathbf{F}_{s,m,n}(r, \theta, \phi)$ is the well-defined spherical wave function and the general form is in (7), [37]. In this formula, \hat{r} , $\hat{\theta}$ and $\hat{\phi}$ are unit vectors. $z_n(\kappa r)$ is the spherical Hankel function of the first kind, and the term $P_n^{(m)}(\cos \theta)$ in (7) is the normalized associated Legendre function. For TM waves, the magnetic field is described by \mathbf{F}_{1mn} , and \mathbf{F}_{2mn} represents the corresponding electric fields. For TE waves, this is interchanged.

The electric field can be decomposed into:

$$\mathbf{e}(r, \theta, \phi) = \left[E^{\hat{r}}(r, \theta, \phi), E^{\hat{\theta}}(r, \theta, \phi), E^{\hat{\phi}}(r, \theta, \phi) \right], \quad (6)$$

with the superscripts \hat{r} , $\hat{\theta}$ and $\hat{\phi}$ representing r -component, θ -component, and ϕ -component, respectively.

The spherical wave expansion can be interpreted as the decomposition of an electromagnetic field into a series of orthogonal components with different basic modes. To determine the spherical mode coefficients $Q_{s,m,n}$, we can

$$\mathbf{F}_{mn}(r, \theta, \phi) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{n(n+1)}} \left(-\frac{m}{|m|} \right)^m z_n(\kappa r) P_n^{(m)}(\cos \theta) e^{-jm\phi}, \quad (7a)$$

$$\begin{aligned} \mathbf{F}_{1mn}(r, \theta, \phi) &= \nabla \mathbf{F}_{mn}(r, \theta, \phi) \times \hat{r} \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{n(n+1)}} \left(-\frac{m}{|m|} \right)^m \left\{ z_n(\kappa r) \frac{im P_n^{(m)}(\cos \theta)}{\sin \theta} e^{-jm\phi} \hat{\theta} - z_n(\kappa r) \frac{-jm P_n^{(m)}(\cos \theta)}{\sin \theta} e^{-jm\phi} \hat{\phi} \right\}, \end{aligned} \quad (7b)$$

$$\mathbf{F}_{2mn} = \kappa^{-1} \nabla \times \mathbf{F}_{1mn}(r, \theta, \phi). \quad (7c)$$

reformulate the expansion (5) into matrix form

$$\mathbf{E} = \kappa\sqrt{\eta}\mathbf{F}\mathbf{Q}, \quad (8)$$

where $\mathbf{E} \in \mathbb{C}^{3P \times N_T}$, $\mathbf{F} \in \mathbb{C}^{3P \times 2N(N+2)}$, and $\mathbf{Q} \in \mathbb{C}^{2N(N+2) \times N_T}$ represent the spatially sampled electric field matrix, the basic spherical mode matrix, and the spherical mode coefficient matrix, respectively, with P being the number of spatial sampling points. Their specific representations for the n th antenna are as follows [40]

$$\mathbf{E}(:, n) = [E_n^{\hat{r}}(r_1, \theta_1, \phi_1), E_n^{\hat{\theta}}(r_1, \theta_1, \phi_1), E_n^{\hat{\phi}}(r_1, \theta_1, \phi_1), \dots, E_n^{\hat{r}}(r_P, \theta_P, \phi_P), E_n^{\hat{\theta}}(r_P, \theta_P, \phi_P), E_n^{\hat{\phi}}(r_P, \theta_P, \phi_P)]^T, \quad (9)$$

$$\mathbf{Q}(:, n) = [Q_{1,-1,1}^n, Q_{2,-1,1}^n, \dots, Q_{2,N,N}^n]^T, \quad (10)$$

$$\mathbf{F} = \begin{bmatrix} F_{1,1,1}^{\hat{r}}(r_1, \theta_1, \phi_1) & \dots & F_{2,N,N}^{\hat{r}}(r_1, \theta_1, \phi_1) \\ F_{1,1,1}^{\hat{\theta}}(r_1, \theta_1, \phi_1) & \dots & F_{2,N,N}^{\hat{\theta}}(r_1, \theta_1, \phi_1) \\ F_{1,1,1}^{\hat{\phi}}(r_1, \theta_1, \phi_1) & \dots & F_{2,N,N}^{\hat{\phi}}(r_1, \theta_1, \phi_1) \\ \vdots & \vdots & \vdots \\ F_{1,1,1}^{\hat{r}}(r_P, \theta_P, \phi_P) & \dots & F_{2,N,N}^{\hat{r}}(r_P, \theta_P, \phi_P) \\ F_{1,1,1}^{\hat{\theta}}(r_P, \theta_P, \phi_P) & \dots & F_{2,N,N}^{\hat{\theta}}(r_P, \theta_P, \phi_P) \\ F_{1,1,1}^{\hat{\phi}}(r_P, \theta_P, \phi_P) & \dots & F_{2,N,N}^{\hat{\phi}}(r_P, \theta_P, \phi_P) \end{bmatrix}, \quad (11)$$

where $F_{s,m,n}^{\hat{r}}(r_n, \theta_n, \phi_n)$, $F_{s,m,n}^{\hat{\theta}}(r_n, \theta_n, \phi_n)$ and $F_{s,m,n}^{\hat{\phi}}(r_n, \theta_n, \phi_n)$ represent different orthogonal components of the spherical wave function. Based on the above matrix representation, we can derive the spherical mode coefficients in (10) as

$$\mathbf{Q} = \frac{1}{\kappa\sqrt{\eta}}(\mathbf{F})^\dagger \mathbf{E}. \quad (12)$$

After obtaining the full-wave simulation electromagnetic fields for both the coupled and uncoupled cases, the mutual coupling matrix corresponding to the holographic antenna array can be represented by

$$\mathbf{C} = \mathbf{Q}_C^T(\mathbf{Q}_S^T)^\dagger, \quad (13)$$

where \mathbf{Q}_C is the spherical mode coefficient matrix for coupled cases and \mathbf{Q}_S is the uncoupled one. In summary, we investigate the effects of mutual coupling by regarding the actual radiation pattern of each antenna element as a superposition of the ideal radiation patterns of all elements in the array. The mutual coupling coefficients serve as the weighting coefficients for this superposition. Note that the mutual coupling coefficients are computed based on radiation fields, and therefore the computed coefficients are at the field level.

The impact of mutual coupling is illustrated in Fig. 2. The first two subfigures depict the spatial radiation patterns of an isolated antenna element and a coupled one. Fig. 2(c) displays the spatial superdirective beam achieved by leveraging mutual coupling, whereas Fig. 2(d) shows the spatial beam using a conventional beamforming algorithm. It is evident that neglecting mutual coupling leads to a significant reduction in directivity.

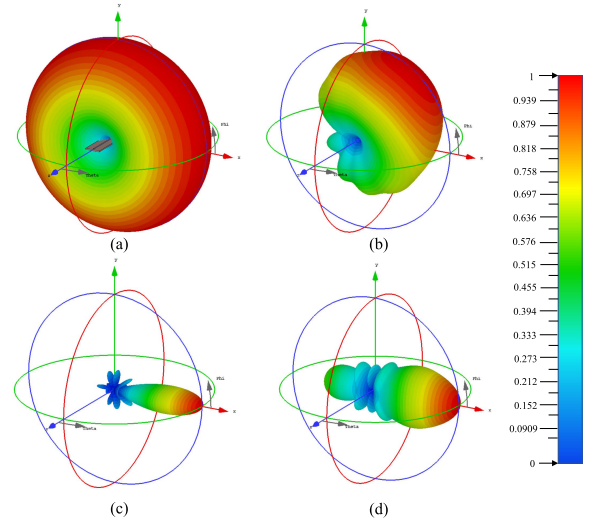


Fig. 2. Radiation pattern for (a) An isolated antenna; (b) One antenna in holographic antenna array with 0.2λ spacing; (c) Spatial beam based on mutual coupling; (d) Spatial beam without considering mutual coupling.

C. Electromagnetic Channel Model

In this subsection, we incorporate the above electromagnetic characteristics of holographic antenna array into a conventional geometrically based stochastic channel model (GSCM) to take both the stochastic characteristics of wireless environment and the spatial radiation pattern of the transmitter into consideration.

Assuming that there are L scatters and K users in the far field region of interest (i.e., we omit variable \mathbf{r} in the radiation pattern function (2)), the classical Saleh-Valenzuela channel model for a single-antenna user k can be written as [41], [42]

$$\mathbf{h}_k = \sum_{l=1}^L \sqrt{P_{lk}} \alpha_{lk} \exp(-j2\pi f \tau_{lk}) \mathbf{C} \mathbf{i} \odot \mathbf{g}(\boldsymbol{\theta}, \boldsymbol{\phi}) \delta(\Omega - \Omega_l), \quad (14)$$

where τ_{lk} , α_{lk} and P_{lk} represent the propagation time delay, channel coefficient and power of the l th path, respectively. Ω_l represents the solid angle of the l th scatterer relative to the transmitter. With $s_{lk} = \sqrt{P_{lk}} \alpha_{lk} \exp(-j2\pi f \tau_{lk})$, the received signal for user k can be formulated as

$$\begin{aligned} y_k &= \int \mathbf{h}_k^T \mathbf{x} d\Omega = \int \sum_{l=1}^L s_{lk} \delta(\Omega - \Omega_l) \mathbf{1}^T \mathbf{x}^{(c)} d\Omega + n \\ &= \sum_{l=1}^L s_{lk} \mathbf{1}^T (\mathbf{C} \mathbf{i} \odot \mathbf{g}(\Omega_l) \odot \mathbf{x}) + n \\ &= \sum_{l=1}^L s_{lk} \mathbf{x}^T (\mathbf{C} \hat{\mathbf{I}} \mathbf{g}(\Omega_l)) + n, \end{aligned} \quad (15)$$

where $\mathbf{1} \in \mathbb{C}^{N_T \times 1}$, $\hat{\mathbf{I}} = \text{diag}(\mathbf{i})$, and n represent an all-one vector, a diagonal matrix representing analog antenna excitations, and the Gaussian noise term, respectively. Based on (15), the received signal can be written in the matrix form

$$\mathbf{y} = (\mathbf{C} \hat{\mathbf{I}} \mathbf{G} \mathbf{S})^T \mathbf{x} + \mathbf{n}, \quad (16)$$

where $\mathbf{S} \in \mathbb{C}^{L \times K}$, $\mathbf{G} \in \mathbb{C}^{N_T \times L}$, and \mathbf{n} are the channel coefficient matrix, the spatial sample of the radiation pattern, and an i.i.d. Gaussian noise vector with element variance σ^2 , respectively. \mathbf{G} describes the spatial radiation pattern corresponding to the L scatterers and can be obtained after acquiring the spatial coordinates of scatterers in the environment through sensing or channel estimation. In practical systems, channel state information at the receiver can be obtained via training and subsequently shared with the transmitter via feedback. Techniques for efficient channel estimation leveraging the geometric nature of wireless environments can be found in literature [43], [44], [45], [46], [47], [48], as well as deep learning-based approaches [49], [50], [51], [52], [53] or beam scanning techniques [54], [55], [56]. The derived electromagnetic GSCM can be written as

$$\mathbf{H} = (\mathbf{C}\hat{\mathbf{I}}\mathbf{G}\mathbf{S})^T \in \mathbb{C}^{K \times N_T}, \quad (17)$$

D. Holographic Antenna Array Radiation Efficiency

Besides the impact of distorted radiation patterns on system performance, another significant factor affecting the realized communication rate is antenna radiation efficiency. Based on existing research on superdirective antenna arrays [57], [58], it has been observed that as the interval space between antenna elements decreases, the achievable directivity increases. However, corresponding radiation efficiency decreases, resulting in a potential deterioration in performance due to the decrease in the actual realized antenna gain. Therefore, in practical terms, a denser antenna array does not necessarily guarantee better system performance. The definition of antenna efficiency is given as follows

$$e_{\text{rad}} = \frac{P_{\text{rad}}}{P_{\text{acc}}} = \frac{G(\theta, \phi)}{D(\theta, \phi)}, \quad (18)$$

where e_{rad} , P_{rad} , and P_{acc} are radiation efficiency, the radiated power, and the accepted power of the antenna, respectively. $D(\theta, \phi)$ and $G(\theta, \phi)$ are defined as the directivity and gain of the antenna, and their specific definitions are provided as follows

$$\begin{aligned} D(\theta, \phi) &= 4\pi \cdot \frac{U}{P_{\text{rad}}}, \\ G(\theta, \phi) &= 4\pi \cdot \frac{U}{P_{\text{rad}}} \frac{R_{\text{rad}}}{R_{\text{rad}} + R_{\text{dis}}}, \end{aligned} \quad (19)$$

where R_{rad} , R_{dis} , and $U = r^2 W_{\text{rad}}$ represent the radiation resistance, dissipation resistance, and radiation intensity (Watt/unit solid angle), respectively.

Since we achieve superdirectivity by leveraging the mutual coupling effect of the holographic array transmitter, a smaller element spacing is preferred for a stronger coupling and more directive spatial beams to increase the concentration of radiated energy. On the other hand, holographic antenna arrays with smaller space between elements tend to have larger dissipation resistances, resulting in a lower radiation efficiency. Therefore, there exists a tradeoff between the radiation efficiency and the directive gain since the actual realized array gain is the product of them (i.e., $G(\theta, \phi) = e_{\text{rad}} \cdot D(\theta, \phi)$), which means that there is an optimal array geometry, and

our following simulation results also further validate this conjecture.

III. PROPOSED EHB SCHEME

To leverage superdirectivity and programmable array radiation pattern, the EHB scheme is introduced in this section. We consider the antenna radiation efficiency to indicate the received signal power in practical scenario. Then, the 3D holographic array architecture and the electromagnetic beam-forming algorithm will be provided in the sequel.

A. 3D Holographic Antenna Array Structure

The directivity expression for an antenna array can be written as follows in the absence of mutual coupling:

$$D = \frac{\mathbf{i}^H |g(\theta, \phi)|^2 \mathbf{e}(\theta, \phi) \mathbf{e}^H(\theta, \phi) \mathbf{i}}{\mathbf{i}^H \left(\int_0^{2\pi} \int_0^\pi |g(\theta, \phi)|^2 \mathbf{e}(\theta, \phi) \mathbf{e}^H(\theta, \phi) \sin(\theta) d\theta d\phi \right) \mathbf{i}}, \quad (20)$$

where $\mathbf{e} = [e^{-j\kappa \hat{\mathbf{r}} \cdot \mathbf{r}_1}, e^{-j\kappa \hat{\mathbf{r}} \cdot \mathbf{r}_2}, \dots, e^{-j\kappa \hat{\mathbf{r}} \cdot \mathbf{r}_{N_T}}]^T$ represents the array steering vector, with $\hat{\mathbf{r}}$ being the unit vector in direction (θ, ϕ) and \mathbf{r}_n being the Cartesian coordinates of the n th antenna. To maximize the achievable directivity, the optimal \mathbf{i} can be derived as:

$$\mathbf{i}_{\text{opt}} = \gamma \mathbf{Z}^{-1} \mathbf{e}(\theta, \phi), \quad (21)$$

with $\mathbf{Z} = \frac{\int_0^{2\pi} \int_0^\pi |g(\theta, \phi)|^2 \mathbf{e}(\theta, \phi) \mathbf{e}^H(\theta, \phi) \sin(\theta) d\theta d\phi}{\int_0^{2\pi} \int_0^\pi |g(\theta, \phi)|^2 \sin(\theta) d\theta d\phi}$ denoted as the impedance matrix and γ denoting the power normalization factor. For lossless antennas, the radiation power P_{rad} equals the input power P_{in} , and the maximum array gain can be written as follows [12]:

$$A = \max \frac{P_{\text{Rx}}}{P_{\text{Rx}}|_{N_T=1}} = N \frac{\mathbf{e}^H(\theta, \phi) \mathbf{Z}^{-1} \mathbf{e}(\theta, \phi)}{\mathbf{e}^H(\theta, \phi) \mathbf{e}(\theta, \phi)}, \quad (22)$$

where P_{Rx} represents the received power. Specifically, the array gain depends on the the beamforming direction and array geometry (via \mathbf{Z}). For isotropic elements along the z axis, the largest directivity is obtained in the direction $\theta = 0$ (end-fire), and approaches N_T^2 , for small antenna separation.

Inspired by the former conclusion, we find that the superdirectivity origins from the dense element arrangements along a specific direction and the corresponding optimal excitation currents. Thus, the \mathbf{Z} matrix of a densely placed 3D holographic antenna array would fulfill the geometry requirements of superdirectivity in arbitrary directions. Considering the mutual coupling effect of holographic antenna array on actual radiation pattern, the optimal excitation currents in (21) can be modulated as:

$$\mathbf{i}_{\text{opt}}^{(c)} = \gamma \mathbf{C}^{-1} \mathbf{Z}^{-1} \mathbf{e}(\theta, \phi), \quad (23)$$

and multiple superdirectivity beams can be implemented through a linear combination of excitation current vectors in different directions as $\mathbf{i}_{\text{opt}}^{(c)} = \mathbf{i}_{\text{opt},1}^{(c)} + \mathbf{i}_{\text{opt},2}^{(c)}$. As an example, Fig. 3 shows a typical 3D geometry with multiple layers of planar aperture. Moreover, two superdirective spatial beams in orthogonal directions (broadside and end-fire directions) are realized,

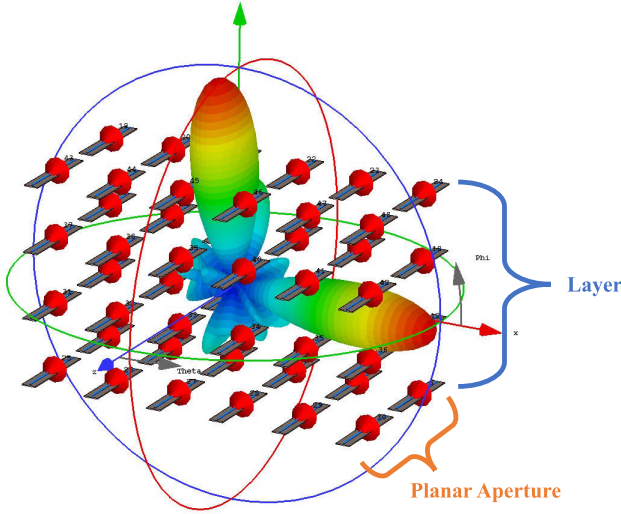


Fig. 3. Multiple superdirective beams of 3D holographic antenna array.

demonstrating the effectiveness of our proposed structure and design principle. Therefore, the real-time optimization technique of circuit parameters can be utilized to provide an approximately coincident array gain in any desired direction and eliminate the edge pattern distortion.

B. Proposed EHB Algorithm

Based on the electromagnetic system model and the 3D antenna array structure that we have introduced, we aim to fully leverage the programmable electromagnetic characteristics of the holographic antenna array. To achieve this, we introduce EHB to modulate the radiation pattern of the array and mitigate radiation distortion. Specifically, the characteristics of our proposed EHB algorithm are twofold: i) Electromagnetic analog precoding method: By real-time programming the circuit parameters, including parameters of programmable antennas and the coupling between antennas, load impedance or excitation currents of the transmitter, this approach can realize the real-time adjustment of the radiation pattern to achieve an approximately coincide gain for users in various directions. ii) Electromagnetic digital precoding method: Its objective is to obtain the optimal match filter to further improve the spectral efficiency of communication system based on the channel information after the analog precoding. Then, analog precoding and digital precoding are iteratively optimized until they both achieve the optimum performance.

Next we show the design details of the analog excitation vector \mathbf{i} and the digital precoding matrix \mathbf{W}_{BB} to optimize overall system performance based on the electromagnetic GSCM derived in Section II.

1) *Single-User Case*: First we consider single-user communication scenario. Referring to the general formulas (16), the received signal in a single-user communication scheme can be formulated as

$$\begin{aligned} y &= \sqrt{e_{\text{rad}}} \mathbf{h}^T \mathbf{x} + n \\ &= \sqrt{e_{\text{rad}}} (\mathbf{C} \hat{\mathbf{I}} \mathbf{G} \mathbf{s})^T \mathbf{x} + n \\ &= \sqrt{e_{\text{rad}}} (\mathbf{C} \hat{\mathbf{I}} \mathbf{G} \mathbf{s})^T \mathbf{w}_{BB} t + n, \end{aligned} \quad (24)$$

and the corresponding received signal-to-noise-ratio (SNR) can be represented as

$$\gamma = \frac{e_{\text{rad}} |(\mathbf{C} \hat{\mathbf{I}} \mathbf{G} \mathbf{s})^T \mathbf{w}_{BB}|^2}{\sigma^2}, \quad (25)$$

where $\mathbf{w}_{BB} \in \mathbb{C}^{N_T \times 1}$ represents the digital beamforming vector and n is the noise term following Gaussian random distribution $\mathcal{N}(0, \sigma^2)$. The transmit symbol is normalized such that $|t|^2 = 1$. Then the corresponding EHB optimization problem can be formulated as

$$\begin{aligned} P1 : \max_{\hat{\mathbf{I}}, \mathbf{w}_{BB}} \quad & \gamma \\ \text{s.t.} \quad & \|\hat{\mathbf{I}}\|_F^2 \leq P_{T,\text{analog}}, \|\mathbf{w}_{BB}\|_2^2 \leq P_{T,\text{digital}} \\ & \hat{\mathbf{I}} = \text{diag}(\mathbf{i}) \\ & P_{T,\text{analog}} + P_{T,\text{digital}} \leq P_T, \end{aligned} \quad (26)$$

where P_T , $P_{T,\text{analog}}$, and $P_{T,\text{digital}}$ represent the available transmit power, the analog excitation power, and the digital precoding power, respectively [59]. Additionally, the radiation efficiency, which is mainly determined by the geometry of the 3D holographic array, also impacts the received SNR. However, for a fixed structure, different beamforming directions are not expected to cause a significant variation in radiation efficiency and the mutual coupling matrix.

Therefore, the proposed EHB algorithm for the single-user scheme consists of two main procedures. Firstly, a search is performed on the value of the element spacing to find an optimized 3D antenna array under the given operation frequency. This optimization aims to strike a balance between beamforming gain and radiation efficiency. In practice, this can be achieved by referring to the performance curves of realized gain as shown in the simulation section. Secondly, the optimal solutions of analog and digital precoding vectors are applied to maximize the received SNR.

Initially, we apply the following transformation:

$$\hat{\mathbf{I}} \mathbf{G} \mathbf{s} = \text{diag}(\mathbf{G} \mathbf{s}) \mathbf{i}, \quad (27)$$

by substituting the transformation (27) into (26), we obtain the following formulation:

$$\begin{aligned} P2 : \max_{\mathbf{i}, \mathbf{w}_{BB}} \quad & |\mathbf{i}^T \text{diag}(\mathbf{G} \mathbf{s}) \mathbf{C}^T \mathbf{w}_{BB}|^2 \\ \text{s.t.} \quad & \|\mathbf{i}\|_2^2 \leq P_{T,\text{analog}}, \|\mathbf{w}_{BB}\|_2^2 \leq P_{T,\text{digital}} \\ & P_{T,\text{analog}} + P_{T,\text{digital}} \leq P_T, \end{aligned} \quad (28)$$

By expanding the objective function into conjugate form, we have:

$$|\mathbf{i}^T \text{diag}(\mathbf{G} \mathbf{s}) \mathbf{C}^T \mathbf{w}_{BB}|^2 = \mathbf{i}^T \text{diag}(\mathbf{G} \mathbf{s}) \mathbf{C}^T \mathbf{w}_{BB} \times \mathbf{w}_{BB}^H \mathbf{C}^* \text{diag}(\mathbf{G} \mathbf{s})^* \mathbf{i}^*, \quad (29)$$

which is a quadratic form function for variable \mathbf{i} , and the optimal solution \mathbf{i}_{opt} corresponds to the eigenvectors of the intermediate matrix with respect to the variable \mathbf{i} :

$$\mathbf{i}_{\text{opt}} = \alpha \text{diag}(\mathbf{G} \mathbf{s})^* \mathbf{C}^H \mathbf{w}_{BB}^*, \quad (30)$$

where α is the power normalization factor to satisfy the analog power constraint. Similarly, we can derive the optimal digital precoding vector as

$$\mathbf{w}_{BB,\text{opt}} = \beta \mathbf{C}^* \text{diag}(\mathbf{G} \mathbf{s})^* \mathbf{i}^*, \quad (31)$$

where β is the power normalization factor of digital precoding. By combining (30) and (31), it is easy to prove that

$$\begin{aligned} \frac{1}{\alpha\beta^*} \mathbf{i}_{\text{opt}} &= \text{diag}(\mathbf{G}\mathbf{s})^* \mathbf{C}^H \mathbf{C} \text{diag}(\mathbf{G}\mathbf{s}) \mathbf{i}_{\text{opt}} \\ \frac{1}{\alpha^* \beta} \mathbf{w}_{BB, \text{opt}} &= \mathbf{C}^* \text{diag}(\mathbf{G}\mathbf{s})^* \text{diag}(\mathbf{G}\mathbf{s}) \mathbf{C}^T \mathbf{w}_{BB, \text{opt}}, \end{aligned} \quad (32)$$

by substituting (32), (30) and (31) into $P2$, it can be easily proven that under constant power constraints, the optimal \mathbf{i} and \mathbf{w}_{BB} are the eigenvectors corresponding to the largest eigenvalues of matrices $\text{diag}(\mathbf{G}\mathbf{s})^* \mathbf{C}^H \mathbf{C} \text{diag}(\mathbf{G}\mathbf{s})$ and $\mathbf{C}^* \text{diag}(\mathbf{G}\mathbf{s})^* \text{diag}(\mathbf{G}\mathbf{s}) \mathbf{C}^T$, respectively, and they can be obtained by taking the SVD decomposition.

2) *Multi-User Scheme*: In this subsection, we extend the previous single-user communication scheme to a multi-user scenario. Based on the electromagnetic transmit signal and channel models introduced in section II, the received signal can be expressed as follows:

$$\begin{aligned} \mathbf{y} &= \sqrt{e_{\text{rad}}} \mathbf{H} \mathbf{x} + \mathbf{n} \\ &= \sqrt{e_{\text{rad}}} (\mathbf{C} \hat{\mathbf{I}} \mathbf{G} \mathbf{S})^T \mathbf{x} + \mathbf{n} \\ &= \sqrt{e_{\text{rad}}} (\mathbf{C} \hat{\mathbf{I}} \mathbf{G} \mathbf{S})^T \mathbf{W}_{BB} \mathbf{t} + \mathbf{n}, \end{aligned} \quad (33)$$

From (33), it can be seen that radiation efficiency, mutual coupling matrix \mathbf{C} , and analog beamforming excitation $\hat{\mathbf{I}}$ are the main factors influencing system performance. Then, we can formulate the sum rate maximization problem as

$$\begin{aligned} P3: \max_{\hat{\mathbf{I}}, \mathbf{W}_{BB}} & \sum_{k=1}^K \log(1 + \gamma_k) \\ \text{s.t.} & \|\hat{\mathbf{I}}\|_F^2 \leq P_{T, \text{analog}}, \|\mathbf{W}_{BB}\|_F^2 \leq P_{T, \text{digital}} \\ & \hat{\mathbf{I}} = \text{diag}(\mathbf{i}) \\ & P_{T, \text{analog}} + P_{T, \text{digital}} \leq P_T, \end{aligned} \quad (34)$$

and the received signal-to-noise-plus-interference ratio (SINR) for user k can be written as follows due to the mutual interference between different users

$$\gamma_k = \frac{e_{\text{rad}} |[(\mathbf{C} \hat{\mathbf{I}} \mathbf{G} \mathbf{S})(:, k)]^T \mathbf{W}_{BB}(:, k)|^2}{e_{\text{rad}} \sum_{i \neq k} |[(\mathbf{C} \hat{\mathbf{I}} \mathbf{G} \mathbf{S})(:, i)]^T \mathbf{W}_{BB}(:, i)|^2 + \sigma^2}, \quad (35)$$

where $\mathbf{A}(:, k)$ represents the k th column of matrix \mathbf{A} .

It can be seen that Eq. (35) is a classical multi-user digital precoding problem. However, for analog precoding design, it is still challenging to directly optimize the precoding matrix \mathbf{I} based on (35). Therefore, we reformulate (35) and propose a unified optimization procedure for both analog and digital precodings. By defining $\mathbf{X} = \mathbf{G}\mathbf{S} \in \mathbb{C}^{N_T \times K}$, we can rewrite the k th column of the matrix $\mathbf{C} \hat{\mathbf{I}} \mathbf{G} \mathbf{S}$ as:

$$\begin{aligned} (\mathbf{C} \hat{\mathbf{I}} \mathbf{G} \mathbf{S})(:, k) &= \mathbf{C} \times \begin{bmatrix} i_1 & 0 & \dots \\ 0 & \ddots & \vdots \\ \vdots & \dots & i_{N_T} \end{bmatrix} \times \begin{bmatrix} x_{1k} \\ x_{2k} \\ \vdots \\ x_{N_T k} \end{bmatrix} \\ &= \mathbf{C} \times \begin{bmatrix} x_{1k} & 0 & \dots \\ 0 & \ddots & \vdots \\ \vdots & \dots & x_{N_T k} \end{bmatrix} \times \begin{bmatrix} i_{1k} \\ i_{2k} \\ \vdots \\ i_{N_T k} \end{bmatrix} \end{aligned}$$

$$= \mathbf{C} \text{diag}(\mathbf{X}(:, k)) \mathbf{i}, \quad (36)$$

where x_{nk} represents the (n, k) th element of matrix \mathbf{X} . Leveraging (36), the SINR of user k can be written as

$$\gamma_k = \frac{e_{\text{rad}} |\mathbf{i}^T \text{diag}(\mathbf{X}(:, k)) \mathbf{C}^T \mathbf{W}_{BB}(:, k)|^2}{e_{\text{rad}} \sum_{i \neq k} |\mathbf{i}^T \text{diag}(\mathbf{X}(:, i)) \mathbf{C}^T \mathbf{W}_{BB}(:, i)|^2 + \sigma^2} \quad (37)$$

Then, the sum rate maximization problem for multi-user EHB can be reformulated as

$$\begin{aligned} P4: \max_{\mathbf{i}, \mathbf{W}_{BB}} & \sum_{k=1}^K \log(1 + \gamma_k) \\ \text{s.t.} & \|\mathbf{i}\|_2^2 \leq P_{T, \text{analog}}, \|\mathbf{W}_{BB}\|_F^2 \leq P_{T, \text{digital}} \\ & P_{T, \text{analog}} + P_{T, \text{digital}} \leq P_T, \end{aligned} \quad (38)$$

which is a non-convex optimization problem and it is also difficult to directly solve it. To address this, we employ the AO framework that iteratively optimizes the electromagnetic analog and digital precodings. Specifically, an SDR technique is employed to solve the auxiliary matrix \mathbf{R} of analog excitations \mathbf{i} , while auxiliary matrices \mathbf{F}_k are used to optimize the precoding matrices (digital part) [60], [61]. The solution details of EHB is outlined as follows:

Firstly, by introducing auxiliary variables t_k and u_k , we can reformulate $P4$ to optimize the electromagnetic digital beamforming when fixing the analog part

$$P5: \max_{\mathbf{F}_k} \sum_{k=1}^K \log(1 + t_k) \quad (39a)$$

$$\text{s.t.} \quad u_k \geq \sum_{i \neq k} \text{Tr}(\mathbf{H}_k \mathbf{F}_i) + \sigma^2 \quad (39b)$$

$$\frac{t_k^{(n)}}{2u_k^{(n)}} u_k^2 + \frac{u_k^{(n)}}{2t_k^{(n)}} t_k^2 \leq \text{Tr}(\mathbf{H}_k \mathbf{F}_k) \quad (39c)$$

$$\sum_{k=1}^K \text{Tr}(\mathbf{F}_k) \leq P_{T, \text{digital}} \quad (39d)$$

$$\text{rank}(\mathbf{F}_k) = 1, \mathbf{F}_k \succeq \mathbf{0}, \quad (39e)$$

where the superscript (n) is the iteration index, $\mathbf{F}_k = \mathbf{W}_{BB}(:, k) \mathbf{W}_{BB}^H(:, k)$ and $\mathbf{H}_k = \mathbf{C}^* \text{diag}(\mathbf{X}^*(:, k)) \mathbf{i}^* \mathbf{i}^T \text{diag}(\mathbf{X}(:, k)) \mathbf{C}^T$.

Secondly, by fixing the digital beamforming and introducing similar auxiliary variables, we can reformulate $P4$ similarly as above to optimize the electromagnetic analog beamforming as follows:

$$P6: \max_{\mathbf{R}} \sum_{k=1}^K \log(1 + t_k) \quad (40a)$$

$$\text{s.t.} \quad u_k \geq \sum_{i \neq k} \text{Tr}(\mathbf{R} \mathbf{Q}_i) + \sigma^2 \quad (40b)$$

$$\frac{t_k^{(n)}}{2u_k^{(n)}} u_k^2 + \frac{u_k^{(n)}}{2t_k^{(n)}} t_k^2 \leq \text{Tr}(\mathbf{R} \mathbf{Q}_k) \quad (40c)$$

$$\text{Tr}(\mathbf{R}) \leq P_{T, \text{analog}} \quad (40d)$$

$$\text{rank}(\mathbf{R}) = 1, \mathbf{R} \succeq \mathbf{0}, \quad (40e)$$

where $\mathbf{R} = \mathbf{i}^* \mathbf{i}^T$ and $\mathbf{Q}_i = \text{diag}(\mathbf{X}(:, i)) \mathbf{C}^T \mathbf{W}_{BB}(:, i) \mathbf{W}_{BB}^H(:, i) \mathbf{C}^* \text{diag}(\mathbf{X}^*(:, i))$.

Finally, after dropping the non-convex rank-one constraint, the two optimization problems mentioned above both can be reformulated as SDR problems, and solved alternatively by optimizing the digital precoding matrices and analog excitation vectors using convex solvers such as CVX. Note that the SDR problem is a convex optimization problem, and each iteration's solution is the optimal. Therefore, iteratively solving the optimization problems and updating the variables will either increase or at least maintain the value of the objective function. Given limited transmit power, the proposed iterative algorithm guarantees that the obtained sum rate value forms a monotonically non-decreasing sequence with an upper bound. The convergence condition for Algorithm 1 is achieved in terms of the rate value in two successive iterations with variations less than a specified threshold, denoted as ϵ .

Now we reconsider the rank-one constraint. As mentioned above, to formulate SDR problems, we removed the rank-one constraint in the above two optimization procedures. Fortunately, it can be proven that the solutions for $P5$ and $P6$ also satisfy the rank-one constraint.

Theorem 1: The obtained auxiliary digital precoding matrix \mathbf{F}_k and auxiliary analog precoding matrix \mathbf{G} in $P5$ and $P6$ satisfy $\text{rank}(\mathbf{F}_k) = 1$, $\text{rank}(\mathbf{G}) = 1$.

Proof: Please see Appendix. \square

Algorithm 1 EHB Scheme for 3D Holographic Array

Input: Number of antennas N_T , user number K , frequency f , antenna element radiation pattern $g(r, \theta, \phi)$, geometrical channel coefficients s_{lk} , scatter angular information Ω_l and mutual coupling matrix \mathbf{C}

Output: Electromagnetic analog precoding vector \mathbf{i} and digital precoding matrix \mathbf{W}_{BB}

```

1: for  $l = 1, 2, \dots, L$  do
2:   Compute the EM channel vector  $\mathbf{h}_k$  by (14);
3: end for
4: Construct the EM channel matrix  $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_K]^T$ ;
5: Initialize  $\mathbf{i} = \mu_i \mathbf{C}^{-1} \mathbf{Z}^{-1} \mathbf{e}$  and  $\mathbf{W}_{BB} = \mu_w \frac{\mathbf{H}_H^w}{(\mathbf{H}_w \mathbf{H}_w^H)^{-1}}$ ,
   with  $\mathbf{H}_w = [\mathbf{C} \text{diag}(\mathbf{i}) \mathbf{G} \mathbf{S}]^T$ ;
6: for  $i = 1, 2, \dots, N_{iter}$  do
7:   Determine the auxiliary rank-one matrices  $\mathbf{F}_k$  for
   digital precoding matrix optimization based on (39);
8:   Determine the auxiliary rank-one matrix  $\mathbf{R}$  for antenna
   excitation vector optimization based on (40);
9: end for
10: for  $k = 1, 2, \dots, K$  do
11:   Compute  $\mathbf{f}_k$  through SVD of matrix  $\mathbf{F}_k$ ;
12: end for
13: Construct  $\mathbf{W}_{BB} = [\mathbf{f}_1, \dots, \mathbf{f}_K]$ ;
14: Obtain  $\mathbf{i}$  through SVD of  $\mathbf{R}$ ;
15: return  $\mathbf{i}$  and  $\mathbf{W}_{BB}$ 

```

The employed EHB algorithm is summarized in **Algorithm 1**. We first compute the electromagnetic channel vector \mathbf{h}_k for each user and construct the overall channel matrix \mathbf{H} based on the estimated GSCM, as shown in lines 1-4. We initialize the analog excitation vector and digital beamforming matrix based on the impedance matrix \mathbf{Z} and

steering vector \mathbf{e} . The initial point is determined by adopting the excitation currents that maximize the array directivity in the absence of mutual coupling, with \mathbf{Z} defined in (21), as shown in line 5. Then, according to the AO optimization based on SDR method in $P5$ and $P6$, we can obtain the optimized rank-one auxiliary matrices, as shown in lines 7-8. By taking the SVD of auxiliary matrices, we can verify the correctness of SDR solutions by checking the rank and obtain the corresponding beamforming vectors, as shown in lines 11 and 14. Finally, the digital beamforming matrix can be obtained by concatenating the derived precoders for each user, as shown in line 13.

C. Complexity Analysis

In this subsection, we analyze the computational complexity of our proposed EHB scheme. Recall that the above convex restrictions in $P5$ and $P6$ involve only linear matrix inequality (LMI) and second-order cone (SOC) constraints. Therefore, they can all be solved by a standard interior-point method (IPM) [62], [63], [64]. This suggests that the worst-case runtime of such a method can be used to compare the computational complexities of the formulations. From [65], the basic elements in the complexity analysis of IPMs can be formulated as follows

$$\min_{\mathbf{z} \in \mathbb{R}^{\mathcal{N}}} \mathbf{c}^T \mathbf{z} \quad (41a)$$

$$\sum_{i=1}^{\mathcal{N}} z_i \mathbf{A}_i^j - \mathbf{B}^j \in \mathbb{S}_+^{k_j} \quad \text{for } j = 1, \dots, q, \quad (41b)$$

$$\mathbf{D}^j \mathbf{z} - \mathbf{b}^j \in \mathbb{L}^{k_j} \quad j = q+1, \dots, p, \quad (41c)$$

where \mathcal{N} is the number of unknowns, $\mathbf{A}_i^j, \mathbf{B}^j \in \mathbb{S}_+^{k_j}$ for $i = 1, \dots, \mathcal{N}$ and $j = 1, \dots, q$; $\mathbf{D}^j \in \mathbb{R}^{k_j \times \mathcal{N}}$ and $\mathbf{b}^j \in \mathbb{R}^{k_j}$ for $j = q+1, \dots, p$; $\mathbf{c} \in \mathbb{R}^{\mathcal{N}}$; \mathbb{S}_+^k is the set of $k \times k$ real positive semidefinite matrices and \mathbb{L}^k is the second-order cone of dimension $k \geq 1$. Specifically, linear constraint $\mathbf{a}^T \mathbf{z} - b \geq 0$ is equivalent to size 1 LMI constraint $\mathbf{a}^T \mathbf{z} - b \in \mathbb{S}_+^1$ and can be put into form (41b).

The complexity of solving (41) through a generic IPM can be decomposed into the following two parts [66]:

1) *Iteration Complexity:* Let $\epsilon > 0$ be the iteration accuracy, the number of iterations required to reach an ϵ -optimal solution of (41) is on the order of $\sqrt{\beta} \cdot \ln(1/\epsilon)$, where $\beta(\Lambda) = \sum_{j=1}^q k_j + 2(p-q)$ is the barrier parameter associated with the cone $\Lambda = \prod_{j=1}^q \mathbb{S}_+^{k_j} \times \prod_{j=q+1}^p \mathbb{L}^{k_j}$. Roughly speaking, the barrier parameter measures the geometric complexity of the conic constraints in (41).

2) *Per-iteration Computation Complexity:* In each iteration, the computation cost is determined by (i) the formation of the $\mathcal{N} \times \mathcal{N}$ coefficient matrix \mathbf{H} of the linear system, and (ii) the factorization of \mathbf{H} . The cost of forming \mathbf{H} is on the order of

$$C_{form} = \mathcal{N} \sum_{j=1}^q k_j^3 + \mathcal{N}^2 \sum_{j=1}^q k_j^2 + \mathcal{N} \sum_{j=q+1}^p k_j^2$$

$$C_{fact} = \mathcal{N}^3, \quad (42)$$

TABLE I
COMPLEXITY ANALYSIS OF P5 AND P6

Method	Complexity Order ($n_1 = \mathcal{O}(KN_T^2)$) ($n_2 = \mathcal{O}(N_T^2)$)
P5	$\sqrt{KN_T + 3K + 1} \cdot n_1 \cdot [(KN_T^3 + 10K + 1) + n_1(KN_T^2 + K + 1) + n_1^2]$
P6	$\sqrt{N_T + 3K + 1} \cdot n_2 \cdot [(N_T^3 + 10K + 1) + n_2(N_T^2 + K + 1) + n_2^2]$

The total computation cost per iteration is on the order of $C_{form} + C_{fact}$, and the corresponding complexity of a generic IPM is on the order of $\sqrt{\beta(\Lambda)} \cdot (C_{form} + C_{fact}) \ln(1/\epsilon)$.

For P5, there are $K + 1$ LMI constraints of size 1, K LMI constraints of size N_T and K second-order cone constraints of size 3. For P6, there are $K + 1$ LMI constraints of size 1, 1 LMI constraints of size N_T and K second-order cone constraints of size 3. Their corresponding complexity orders are summarized in Table. I.

IV. NUMERICAL RESULTS

In this section, simulation results are presented to evaluate the performance of our proposed EHB communication scheme based on 3D holographic antenna arrays. The communication system is assumed to operate at 1.6 GHz. A 3D antenna array consisting of printed dipole antennas is equipped at BS to serve K users, with specific antenna hardware parameters as follows: The dipole antenna is printed on a FR-4 substrate measuring $12.2 \text{ mm} \times 78 \text{ mm}$, and the dimension of a dipole element is $1 \text{ mm} \times 71.5 \text{ mm}$. The digital transmit power constraint is determined based on the SNR settings in simulations while the analog power constraint for antenna excitation is set to 1 W. We consider the area corresponding to four linear half-wavelength spaced antennas as the aperture area of a single layer. Therefore, the number of antennas of one layer is denoted as $N_t = \lfloor \frac{1.5\lambda}{d} \rfloor + 1$, and the total number of antennas is given by $N_T = \nu N_t$, where d and ν represent the element spacing and number of 3D layers, respectively. The convergence parameter for Algorithm 1 is set to $\epsilon = 0.1$. Additionally, we assume that there are L scatterers and K users distributed randomly in the azimuth plane, considering a non-line-of-sight environment.

For the comparison benchmark, we use the current excitation vector corresponding to maximum directivity in the absence of mutual coupling $\mathbf{i} = \mu_i \mathbf{Z}^{-1} \mathbf{e}$ in [16] and the corresponding zero-forcing (ZF) digital precoding matrix. In the subsequent simulations, we compare the performance of our proposed EHB scheme against the conventional beamforming schemes. 1000 sets of channel coefficients are sampled and the numerical performance results are averaged over these samples.

Fig. 4 shows the achieved directivity and realized gain of the 3D holographic antenna array for different transmit azimuth angles. From the figure, we can see that our proposed 3D array achieves a relatively smooth gain over the whole space, which proves the effectiveness of its structure. Additionally, the directivity curves exhibit peaks at 45 and 135 degrees. This

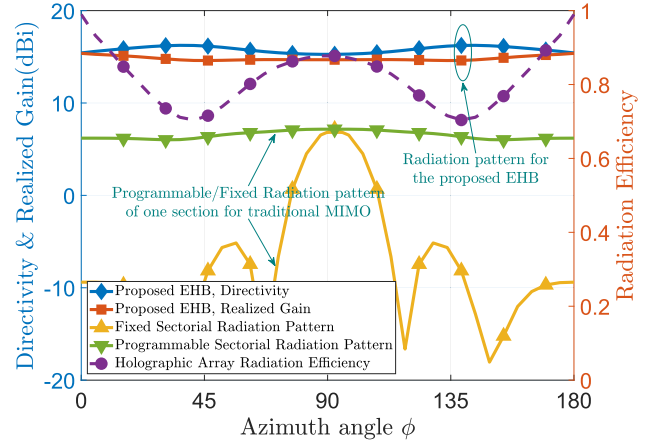


Fig. 4. Achievable directivity and realized gain for EHB system.

phenomenon can be attributed to the fact that these angles are in the vicinity of the diagonal direction of the 3D antenna array. That is to say, from the perspective of beamforming direction, the diagonal direction corresponds to the longest quasi-linear array within the array structure. The presence of these peaks in the directivity curves validates our intuitive design principle.

Fig. 5 shows the rate performance of the EHB scheme under different element spacing and user settings. As can be seen, the ZF precoding without considering the mutual coupling effect achieves the worse rate performance than that of the proposed scheme. Specifically, for the single-user case, the gain of EHB is relatively limited around 10% on average at a 20 dB SNR level. However, in multi-user communication scenarios, exploiting programmable radiation patterns can significantly reduce inter-user interference and improve beam focusing. By leveraging the spatial superdirective beams, the proposed scheme achieves a substantial capacity improvement with an average performance gain of over 150%. Note that the optimal element spacing varies for different 3D antenna structures. This indicates that the geometrical configuration of the 3D holographic antenna array plays a crucial role in achieving the best system performance. The results clearly demonstrate that our proposed scheme outperforms the conventional precoding algorithms in HMIMO communications.

To evaluate the performance of the EHB scheme for different transmission power cases, Fig. 6 shows the sum rate performance under different SNR settings when the antenna spacing is fixed at 0.35 wavelength. Simulation results indicate that larger performance gains are observed as the SNR increases. For comparison, we introduce full-digital precoding and traditional hybrid beamforming in [4] and [67]. Specifically, the 4-layers optimal sum rate gain is achieved by the multi-user 3D antenna array, demonstrating the effectiveness of our proposed EHB algorithm and architecture. The conventional ZF curve tends to saturate because of the mutual interference induced by the mutual coupling effect.

Fig. 7 shows the achievable directivity and corresponding radiation efficiency of the 3D holographic antenna array for different antenna spacings. For comparison, we denote

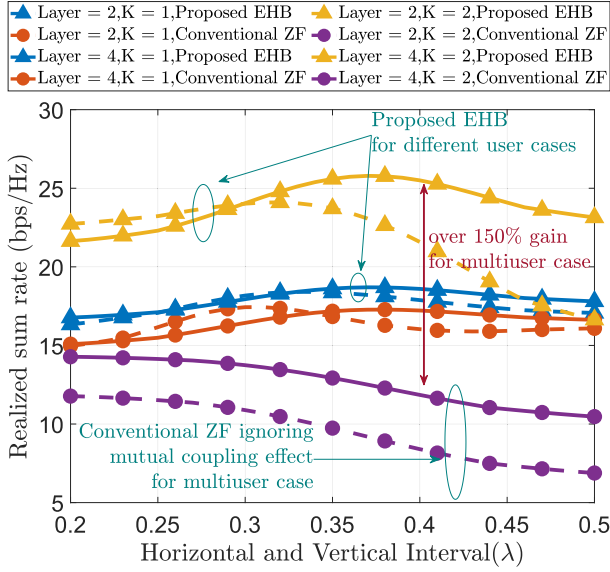


Fig. 5. Achievable rate for EHB system with SNR = 20 dB.

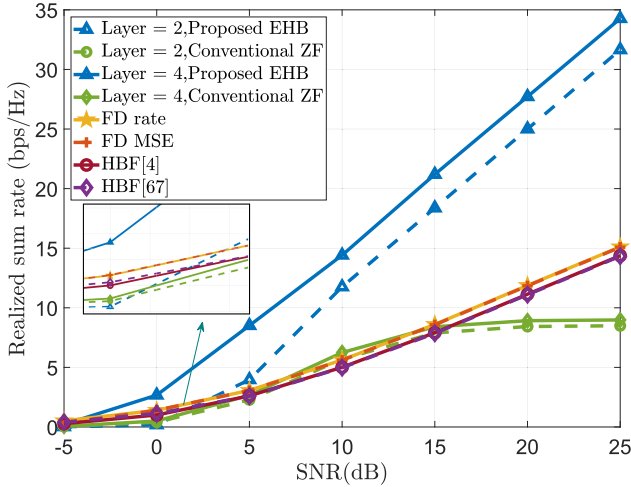


Fig. 6. Achievable sum rate for multi-user EHB system with fixed antenna aperture.

theoretical gain limitation as the realized gain benchmark, expressed as $G = \frac{4\pi A}{\lambda^2}$ [68], where A refers to the array aperture area and λ represents the wavelength. It can be observed that our proposed EHB scheme achieves superdirectivity based on the 3D array architecture while adhering to a fixed aperture area constraint. It is also shown that the radiation efficiency of the 3D array experiences a significant reduction as the antenna spacing decreases, indicating stronger mutual coupling effects. Besides, for a holographic array transmitter with a fixed aperture, when the antenna spacing is smaller than a threshold, for example, 0.3λ in this case, the mutual coupling effect is severe and requires highly accurate excitation current control, which is not achievable under practical hardware constraints, leading to a decreased directivity. The aforementioned optimal antenna spacing in an HMIMO array can vary depending on the specific 3D structure of the array. The array radiation efficiency is obtained through CST simulations.

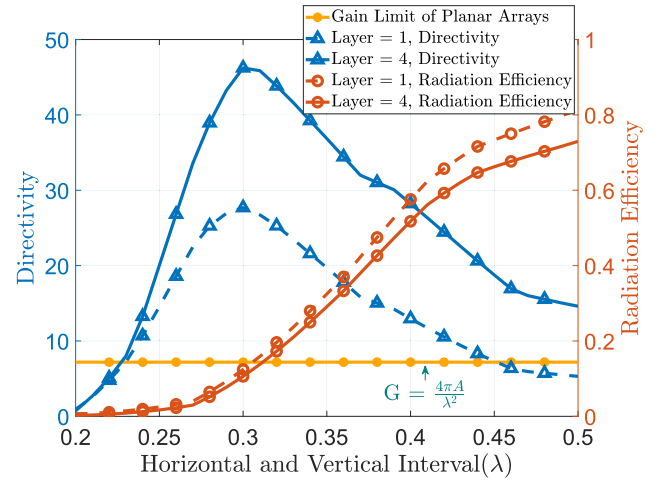


Fig. 7. Achievable directivity for fixed aperture EHB system with varying number of antennas.

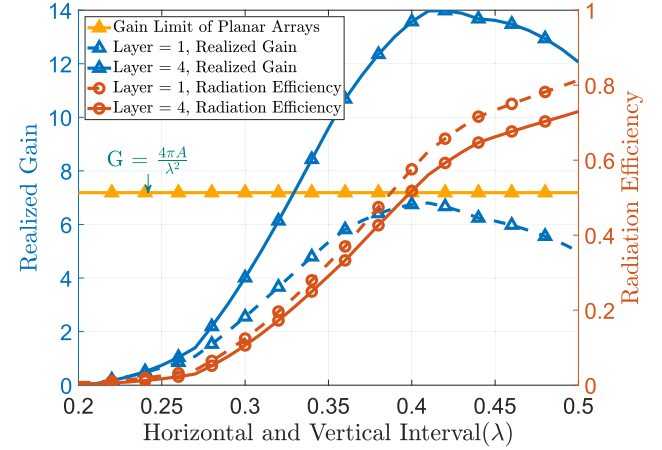


Fig. 8. Achievable realized gain for fixed aperture EHB system with varying number of antennas.

To evaluate the realized gain performance of the 3D array in different antenna spacing cases, in Fig. 8 we show the achievable realized gain and corresponding radiation efficiency of the proposed holographic array when the SNR is fixed at 20 dB. As shown in Fig. 8, the realized gain of planar array approaches the gain limit at 0.4λ through EHB, while our proposed EHB scheme with 3D holographic array structure outperforms the planar gain limitation in regions with relatively high radiation efficiency. However, when the antenna spacing is reduced below a threshold, for example, 0.4λ in this case, the corresponding radiation efficiency decreases rapidly and results in a rapid decrease in realized gain as well. Note that the optimal antenna spacing for maximum gain is different from that of the maximum directivity due to the impact of radiation efficiency.

Figs. 7 and 8 demonstrate the superdirective performance of our proposed EHB scheme based on 3D holographic antenna array within a fixed aperture area. In the following figures, we present simulation results for a fixed antenna number, implying that as the antenna spacing varies, the array aperture area also changes accordingly.

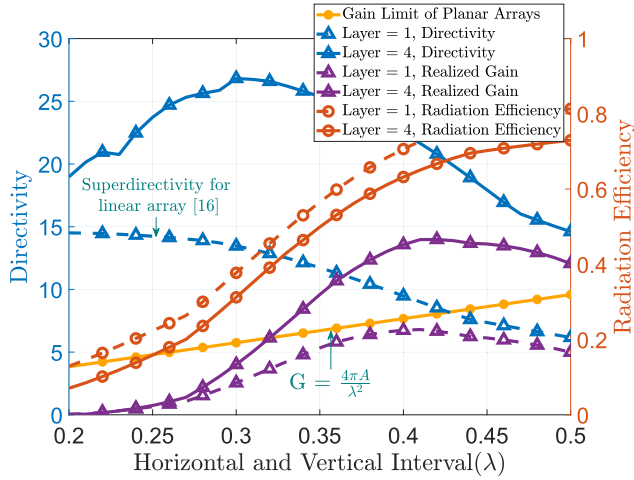


Fig. 9. Achievable directivity for fixed antenna number EHB system with varying array aperture.

In Fig. 9, we present the directivity and realized gain performance of the proposed EHB scheme for 4 antenna element configurations with different antenna spacings while keeping the SNR fixed at 20 dB. It is evident that for planar densely placed linear array, the directivity increases and then gradually stabilizes at N_T^2 as the antenna spacing decreases, which aligns with previous research findings [15], [16]. Furthermore, the realized gain of planar array can also approach the gain limit around 0.4λ spacing. However, note that both the directivity and realized gain achieved by the proposed 3D holographic array surpasses that of the planar linear array, particularly in the region where the antenna element spacing is larger than the aforementioned threshold. This highlights the superior directional and gain performance of the proposed 3D structure compared to its traditional counterpart, emphasizing its potential for practical applications.

At last, we show the achievable communication rate for a fixed antenna number with different spacings in Fig 10. The results further demonstrate that it is necessary to apply EHB to fully utilize the 3D holographic antenna array, otherwise, there is only marginal benefit in expanding the 3D array. This emphasizes the crucial role of the proposed beamforming techniques in harnessing the full potential of the 3D holographic array structure and maximizing the sum rate performance. Additionally, it can be observed that the communication rate exhibits an upward and then downward trend as the antenna spacing increases from 0.2λ to 0.5λ , which is in coherence with the realized gain trend in Fig. 9. The reason for this phenomenon is as follows: in the sub-half-wavelength region, as the antenna spacing increases, the mutual coupling effect decreases, resulting in reduced superdirectivity but higher radiation efficiency. Their tradeoff results in the fluctuation in this region. In the region above the half-wavelength, however, a larger antenna spacing corresponds to a larger aperture area under the constraint of a fixed number of antennas. This leads to a naturally higher array gain and communication rate. In this region, the impact of mutual coupling becomes less significant compared to the increased aperture area, and the fluctuation is due to sidelobes.

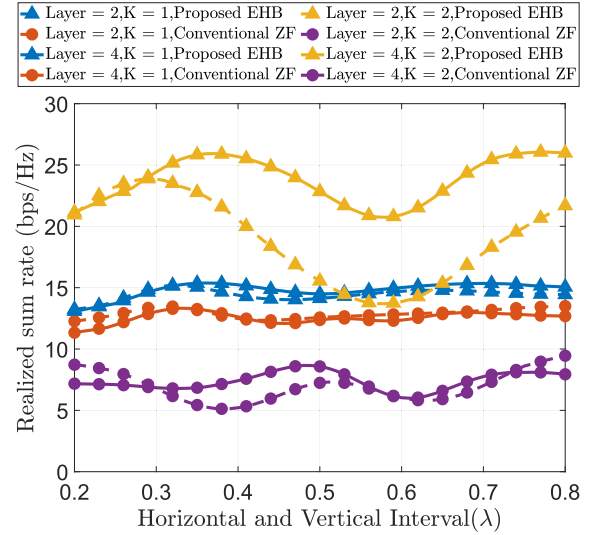


Fig. 10. Achievable sum rate for fixed antenna number with varying array aperture.

V. CONCLUSION

In this paper, we proposed an EHB communication scheme based on 3D holographic antenna arrays that takes the impact of array radiation patterns into account, and incorporates the coupling effect of antenna arrays to have the superdirective gain. Specifically, the implementation of analog precoding involves the real-time adjustment of the radiation pattern to adapt to the wireless environment. Meanwhile, the digital precoding is optimized based on the channel characteristics of analog precoding. As a result, our approach enables the attainment of programmable spatial patterns and a relatively flat beamforming gain, facilitating the realization of multiple superdirective beams in arbitrary directions. Through simulations, we demonstrate the effectiveness of our proposed method. Additionally, we identify the presence of the optimal antenna spacing in the array geometry, considering the trade-off between the radiation efficiency and superdirective gain.

APPENDIX PROOF OF THEOREM 1

Firstly we give the Lagrangian function of (39) without the rank-one constraint as follows:

$$\begin{aligned}
 & F(t_k, b_k, \lambda_1, \lambda_{2,k}, \lambda_{3,k}, \lambda_{4,k}, \mathbf{F}_k) \\
 &= \sum_{k=1}^K \log(1 + t_k) \\
 &+ \lambda_1 \left(P_{T,\text{digital}} - \sum_{k=1}^K \text{Tr}(\mathbf{F}_k) \right) \\
 &+ \sum_{k=1}^K \lambda_{2,k} \left[u_k - \left(\sum_{i \neq k} \text{Tr}(\mathbf{H}_k \mathbf{F}_i) + \sigma^2 \right) \right] \\
 &+ \sum_{k=1}^K \lambda_{3,k} \left[\text{Tr}(\mathbf{H}_k \mathbf{F}_k) - \left(\frac{t_k^{(n)}}{2u_k^{(n)}} u_k^2 + \frac{u_k^{(n)}}{2t_k^{(n)}} t_k^2 \right) \right] \\
 &+ \sum_{k=1}^K \text{Tr}(\lambda_{4,k} \mathbf{F}_k), \tag{43}
 \end{aligned}$$

where $\lambda_1 - \lambda_4$ represent the Lagrange multipliers corresponding to constraints in (39). Since the relaxed SDP problem $P5$ is convex, Slater's condition is satisfied and the gap between the primal problem and its dual problem is zero [69]. Therefore, the Karush-Kuhn-Tucker (KKT) conditions are necessary and sufficient for the optimal solution of $P5$ without the rank-one constraint. The detailed KKT conditions related to the optimal digital beamforming auxiliary matrix \mathbf{F}_k^* can be written as

$$\lambda_1^* \mathbf{I} + \sum_{i \neq k} \lambda_{2,i}^* \mathbf{H}_i - \lambda_{3,k}^* \mathbf{H}_k = \lambda_{4,k}^* \quad (44a)$$

$$\lambda_{4,k}^* \mathbf{F}_k = \mathbf{0} \quad (44b)$$

$$\lambda_{4,k}^* \succeq \mathbf{0}, \quad (44c)$$

where $\lambda_1^* - \lambda_4^*$ are the optimal Lagrange multipliers. Since \mathbf{I} is a full-rank matrix, $\lambda_1^* \geq 0$ and $\lambda_{2,k}^* \geq 0$, we define $\mathbf{Y} = \lambda_1^* \mathbf{I} + \sum_{i \neq k} \lambda_{2,i}^* \mathbf{H}_i$, which is a positive-definite matrix with full rank. Based on this, we have

$$\begin{aligned} \text{rank}(\lambda_{4,k}^*) &= \text{rank}(\mathbf{Y} - \lambda_{3,k}^* \mathbf{H}_k) \\ &\geq \text{rank}(\mathbf{Y}) - \text{rank}(\lambda_{3,k}^* \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H) \\ &\geq \text{rank}(\mathbf{Y}) - 1, \end{aligned} \quad (45)$$

where $\hat{\mathbf{h}}_k = \mathbf{C}^* \text{diag}(\mathbf{X}^*(:, k)) \mathbf{i}^*$. Therefore, we can claim that the rank of $\lambda_{4,k}^*$ is either N_T or $N_T - 1$. If $\text{rank}(\lambda_{4,k}^*) = N_T$, then the optimal \mathbf{F}_k^* should be zero matrix according to (44b), which means that the digital matrix is zero and no signal is transmitted. Thus, the feasible solution should be $\text{rank}(\lambda_{4,k}^*) = N_T - 1$, and the null space of $\lambda_{4,k}^*$ is 1 dimension. From (44b), the derived optimal beamforming auxiliary matrix \mathbf{F}_k^* must lie in the null space of $\lambda_{4,k}^*$, resulting in a rank-one solution, which fulfills the initial constraint we dropped. The same can be proven for $P6$.

REFERENCES

- [1] C. A. Balanis, *Antenna Theory: Analysis and Design*, 4th ed., Hoboken, NJ, USA: Wiley, 2016.
- [2] X. Huang, Y. J. Guo, and J. D. Bunton, "A hybrid adaptive antenna array," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1770–1779, May 2010.
- [3] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.
- [4] 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.
- [5] X. Gao, L. Dai, S. Han, I. Chih-Lin, and R. W. Heath, "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.
- [6] X. Li and A. Alkhateeb, "Deep learning for direct hybrid precoding in millimeter wave massive MIMO systems," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Nov. 2019, pp. 800–805.
- [7] L. Yan, C. Han, and J. Yuan, "A dynamic array-of-subarrays architecture and hybrid precoding algorithms for terahertz wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 9, pp. 2041–2056, Sep. 2020.
- [8] A. Uzkov, "An approach to the problem of optimum directive antenna design," *Comp. Rendus (Doklady) de l'Academie des Sci. de l'URSS*, vol. 53, no. 1, pp. 35–38, 1946.
- [9] M. T. Ivrlac and J. A. Nossek, "The maximum achievable array gain under physical transmit power constraint," in *Proc. Int. Symp. Inf. Theory Appl.*, Dec. 2008, pp. 1–6.
- [10] E. Björnson, L. Sanguinetti, H. Wymeersch, J. Hoydis, and T. L. Marzetta, "Massive MIMO is a reality-what is next five promising research directions for antenna arrays," *Digit. Signal Process.*, vol. 94, pp. 3–20, Nov. 2019.
- [11] A. Bloch, R. G. Medhurst, and S. D. Pool, "A new approach to the design of super-directive aerial arrays," *J. Inst. Electr. Eng.*, vol. 1953, no. 10, pp. 324–326, Oct. 1953.
- [12] M. T. Ivrlac and J. A. Nossek, "High-efficiency super-gain antenna arrays," in *Proc. URSI Int. Symp. Electromagn. Theory*, Aug. 2010, pp. 859–860.
- [13] D. Y. Levin, S. Markovich-Golan, and S. Gannot, "Near-field superdirectivity: An analytical perspective," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 29, pp. 1661–1674, 2021.
- [14] T. L. Marzetta, "Super-directive antenna arrays: Fundamentals and new perspectives," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Nov. 2019, pp. 1–4.
- [15] L. Han, H. Yin, and T. L. Marzetta, "Coupling matrix-based beamforming for superdirective antenna arrays," in *Proc. IEEE Int. Conf. Commun.*, May 2022, pp. 5159–5164.
- [16] L. Han, H. Yin, M. Gao, and J. Xie, "A superdirective beamforming approach with impedance coupling and field coupling for compact antenna arrays," 2023, *arXiv:2302.08203*.
- [17] L. Han and H. Yin, "Superdirectivity-enhanced wireless communications: A multi-user perspective," 2023, *arXiv:2307.06958*.
- [18] C. Huang et al., "Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends," *IEEE Wireless Commun.*, vol. 27, no. 5, pp. 118–125, Oct. 2020.
- [19] E. Björnson et al., "Towards 6G MIMO: Massive spatial multiplexing, dense arrays, and interplay between electromagnetics and processing," 2024, *arXiv:2401.02844*.
- [20] T. Gong et al., "Holographic MIMO communications: Theoretical foundations, enabling technologies, and future directions," *IEEE Commun. Surveys Tuts.*, vol. 26, no. 1, pp. 196–257, 1st Quart., 2024, doi: 10.1109/COMST.2023.3309529.
- [21] A. Pizzo, A. d. J. Torres, L. Sanguinetti, and T. L. Marzetta, "Nyquist sampling and degrees of freedom of electromagnetic fields," *IEEE Trans. Signal Process.*, vol. 70, pp. 3935–3947, 2022.
- [22] R. Ji et al., "Extra DoF of near-field holographic MIMO communications leveraging evanescent waves," *IEEE Wireless Commun. Lett.*, vol. 12, no. 4, pp. 580–584, Apr. 2023.
- [23] X. Gan, C. Huang, Z. Yang, C. Zhong, and Z. Zhang, "Near-field localization for holographic RIS assisted mmWave systems," *IEEE Commun. Lett.*, vol. 27, no. 1, pp. 140–144, Jan. 2023.
- [24] T. Gong et al., "Holographic MIMO communications with arbitrary surface placements: Near-field LoS channel model and capacity limit," 2023, *arXiv:2304.05259*.
- [25] L. Wei et al., "Tri-polarized holographic MIMO surfaces for near-field communications: Channel modeling and precoding design," *IEEE Trans. Wireless Commun.*, vol. 22, no. 12, pp. 8828–8842, Dec. 2023.
- [26] S. S. A. Yuan, X. Chen, C. Huang, and W. E. I. Sha, "Effects of mutual coupling on degree of freedom and antenna efficiency in holographic MIMO communications," *IEEE Open J. Antennas Propag.*, vol. 4, pp. 237–244, 2023.
- [27] H. Zhang et al., "Holographic integrated sensing and communication," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 7, pp. 2114–2130, Jul. 2022.
- [28] X. Zhang, H. Zhang, H. Zhang, and B. Di, "Holographic radar: Target detection enabled by reconfigurable holographic surfaces," *IEEE Commun. Lett.*, vol. 27, no. 1, pp. 332–336, Jan. 2023.
- [29] X. Hu, R. Deng, B. Di, H. Zhang, and L. Song, "Holographic beamforming for ultra massive MIMO with limited radiation amplitudes: How many quantized bits do we need?" *IEEE Commun. Lett.*, vol. 26, no. 6, pp. 1403–1407, Jun. 2022.
- [30] X. Chen, S. Zhang, and Q. Li, "A review of mutual coupling in MIMO systems," *IEEE Access*, vol. 6, pp. 24706–24719, 2018.
- [31] G. Gradoni and M. Di Renzo, "End-to-end mutual coupling aware communication model for reconfigurable intelligent surfaces: An electromagnetic-compliant approach based on mutual impedances," *IEEE Wireless Commun. Lett.*, vol. 10, no. 5, pp. 938–942, May 2021.
- [32] R. Faqiri, C. Saigre-Tardif, G. C. Alexandropoulos, N. Shlezinger, M. F. Imani, and P. del Hougne, "PhysFad: Physics-based end-to-end channel modeling of RIS-parametrized environments with adjustable fading," *IEEE Trans. Wireless Commun.*, vol. 22, no. 1, pp. 580–595, Jan. 2023.
- [33] X. Chen et al., "Revisit to mutual coupling effects on multi-antenna systems," *J. Commun. Inf. Netw.*, vol. 5, no. 4, pp. 411–422, Dec. 2020.

- [34] C. Masouros, M. Sellathurai, and T. Ratnarajah, "Large-scale MIMO transmitters in fixed physical spaces: The effect of transmit correlation and mutual coupling," *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 2794–2804, Jul. 2013.
- [35] P. Wang, M. N. Khormuji, and B. M. Popovic, "Beamforming performances of holographic surfaces," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2022, pp. 43–48.
- [36] S. Sadat, C. Ghobadi, and J. Nourinia, "Mutual coupling compensation in small phased array antennas," in *Proc. IEEE Antennas Propag. Soc. Symp.*, vol. 4, Jun. 2004, pp. 4128–4131.
- [37] R. Cornelius, *Fast Spherical Near-Field Antenna Measurement Methods*. Aachen, Germany: Apprimus Verlag, Jan. 2019.
- [38] L. J. Chu, "Physical limitations of omni-directional antennas," *J. Appl. Phys.*, vol. 19, no. 12, pp. 1163–1175, Dec. 1948.
- [39] R. F. Harrington, "Effect of antenna size on gain, bandwidth, and efficiency," *J. Res. Nat. Bur. Standards, Sect. D: Radio Propag.*, vol. 64D, no. 1, p. 1, Jan. 1960.
- [40] K. Belmkaddem, T. P. Vuong, and L. Rudant, "Analysis of open-slot antenna radiation pattern using spherical wave expansion," *IET Microw. Antennas Propag.*, vol. 9, no. 13, pp. 1407–1411, Oct. 2015.
- [41] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [42] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.
- [43] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, Jun. 2010.
- [44] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *IEEE Int. Symp. Inf. Theory Proc.*, Jul. 2011, pp. 2168–2172.
- [45] C.-K. Wen, S. Jin, K.-K. Wong, J.-C. Chen, and P. Ting, "Channel estimation for massive MIMO using Gaussian-mixture Bayesian learning," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1356–1368, Mar. 2015.
- [46] X. Gan, C. Zhong, C. Huang, Z. Yang, and Z. Zhang, "Multiple RISs assisted cell-free networks with two-timescale CSI: Performance analysis and system design," *IEEE Trans. Commun.*, vol. 70, no. 11, pp. 7696–7710, Nov. 2022.
- [47] Z. Lin, T. Lv, W. Ni, J. A. Zhang, J. Zeng, and R. P. Liu, "Joint estimation of multipath angles and delays for millimeter-wave cylindrical arrays with hybrid front-ends," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4631–4645, Jul. 2021.
- [48] J. Wang et al., "Sparse Bayesian learning-based 3-D radio environment map construction—Sampling optimization, scenario-dependent dictionary construction, and sparse recovery," *IEEE Trans. Cogn. Commun. Netw.*, vol. 10, no. 1, pp. 80–93, Feb. 2024.
- [49] Y. Yang, F. Gao, X. Ma, and S. Zhang, "Deep learning-based channel estimation for doubly selective fading channels," *IEEE Access*, vol. 7, pp. 36579–36589, 2019.
- [50] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Deep learning-based channel estimation for beamspace mmWave massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 852–855, Oct. 2018.
- [51] X. Ru, L. Wei, and Y. Xu, "Model-driven channel estimation for OFDM systems based on image super-resolution network," in *Proc. IEEE 5th Int. Conf. Signal Image Process. (ICSIP)*, Oct. 2020, pp. 804–808.
- [52] C.-J. Chun, J.-M. Kang, and I.-M. Kim, "Deep learning-based channel estimation for massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 4, pp. 1228–1231, Aug. 2019.
- [53] F. Zhu et al., "Beamforming inferring by conditional WGAN-GP for holographic antenna arrays," *IEEE Wireless Commun. Lett.*, vol. 13, no. 7, pp. 2023–2027, Jul. 2024.
- [54] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, Oct. 2013.
- [55] Y. Ding and B. D. Rao, "Dictionary learning-based sparse channel representation and estimation for FDD massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5437–5451, Aug. 2018.
- [56] Ö. T. Demir and E. Björnson, "A new polar-domain dictionary design for the near-field region of extremely large aperture arrays," in *Proc. IEEE 9th Int. Workshop Comput. Adv. Multi-Sensor Adapt. Process. (CAMSAP)*, Dec. 2023, pp. 251–255.
- [57] R. C. Hansen, "Some new calculations on antenna superdirectivity," *Proc. IEEE*, vol. 69, no. 10, pp. 1365–1366, Oct. 1981.
- [58] J. S. McLean, "A re-examination of the fundamental limits on the radiation Q of electrically small antennas," *IEEE Trans. Antennas Propag.*, vol. 44, no. 5, p. 672, May 1996.
- [59] Z. Zhang and L. Dai, "Pattern-division multiplexing for multi-user continuous-aperture MIMO," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 8, pp. 2350–2366, Jun. 2023.
- [60] W. Hao et al., "Robust design for intelligent reflecting surface-assisted MIMO-OFDMA terahertz IoT networks," *IEEE Internet Things J.*, vol. 8, no. 16, pp. 13052–13064, Aug. 2021.
- [61] Z.-Q. Luo, W.-K. Ma, A. M. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.
- [62] R. H. Byrd, M. E. Hribar, and J. Nocedal, "An interior point algorithm for large-scale nonlinear programming," *SIAM J. Optim.*, vol. 9, no. 4, pp. 877–900, Jan. 1999.
- [63] R. Vanderbei and D. Shanno, "An interior-point algorithm for nonconvex nonlinear programming," *Comput. Optim. Appl.*, vol. 13, pp. 231–252, Apr. 1999.
- [64] G. Cornuéjols, J. Peña, and R. Tütüncü, *Optimization Methods in Finance*, 2nd ed., Cambridge, U.K.: Cambridge Univ. Press, 2018.
- [65] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematic, Jan. 2001.
- [66] K.-Y. Wang, A. M. So, T.-H. Chang, W.-K. Ma, and C.-Y. Chi, "Outage constrained robust transmit optimization for multiuser MISO downlinks: Tractable approximations by conic optimization," *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5690–5705, Nov. 2014.
- [67] F. Sotthabhi 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.
- [68] P. Hannan, "The element-gain paradox for a phased-array antenna," *IEEE Trans. Antennas Propag.*, vol. AP-12, no. 4, pp. 423–433, Jul. 1964.
- [69] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, MA, USA: Cambridge Univ. Press, 2004.