Joint Channel Estimation and Data Rate Maximization for Intelligent Reflecting Surface Assisted Terahertz MIMO Communication Systems

XINYING MA[1], ZHI CHEN[1], (Senior Member, IEEE), WENJIE CHEN[1], ZHUOXUN LI[1], YAOJIA CHI[1], CHONG HAN[2], (Member, IEEE), AND SHAOQIAN LI[1], (Fellow, IEEE)

[1]National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China
[2]University of Michigan-Shanghai Jiao Tong University (UM-SJTU) Joint Institute, Shanghai Jiao Tong University, Shanghai 200240, China

Corresponding author: Zhi Chen (chenzhi@uestc.edu.cn)

ABSTRACT Terahertz (THz) communications recently attract significant attention and become an emerging technology pillar for sixth generation (6G) wireless systems. Due to the serious path attenuation of THz signals, THz communication is applicable for the short-distance indoor scenarios. However, the THz waves are easily blocked by obstacles, leading to a communication interruption. To this end, an intelligent reflecting surface (IRS), which interacts with incident THz waves in a controlled manner by adjusting the discrete phase shifts of the IRS elements, is considered as a promising technology to mitigate blockage vulnerability and enhance coverage capability for indoor scenarios. In light of graphene-enabled hardware structure of an IRS, the IRS-assisted THz multiple-input multiple-output (MIMO) system model is developed. Moreover, an iterative atom pruning based subspace pursuit (IAP-SP) scheme is developed for channel estimation. Compared to the classical subspace pursuit (SP) scheme, the proposed IAP-SP algorithm can substantially reduce the computational complexity while maintaining accurate channel recovery. With the estimated channel, a data rate maximization problem is formulated, which can be converted to a discrete phase shift search problem. The exhaustive search method is firstly proposed to obtain the optimal transmission rate but endure extremely high computational burden. Then, a local search method is proposed to decrease the number of possible discrete phase candidates of IRS while undergoes obvious performance loss. Interestingly, a novel feedforward fully connected structure based deep neural network (DNN) scheme is put forward, which has the ability to learn how to output the optimal phase shift configurations by inputting the features of estimated channel. Simulation results demonstrate that, in contrast with the exhaustive search scheme and the local search scheme, the proposed DNN-based scheme achieves a near-optimal communication rate performance. Meanwhile, the DNN-based scheme enormously alleviates the computational complexity and allows for dynamic parameter adaption in rapid-varying channel conditions.

INDEX TERMS Terahertz (THz) communications, sixth generation (6G), intelligent reflecting surface (IRS), channel estimation, deep neural network (DNN).

I. INTRODUCTION

With the explosive growth of the data traffic in diverse communication scenarios, developing new spectrum resources

The associate editor coordinating the review of this manuscript and approving it for publication was Bhaskar Prasad Rimal.

for future sixth generation (6G) communication networks is urgently demanded [1], [2]. To satisfy the bandwidth requirements of the 6G application scenarios, such as holographic communication, precise telemedicine control, broadband local area network (LAN), wireless data center, space communication networks and on-chip communication

[3]–[6], terahertz (THz) band (from 0.1 THz to 10 THz) has drawn more and more attention for 6G wireless networks. Proverbially, THz frequency band bridges the gap between millimeter wave (mmWave) band and optical frequency band. Compared with mmWave communication, THz communication possesses more abundant spectrum resources and is able to provide much higher transmission rates from hundreds of Gigabit-per-second (Gbps) to several Terabit-per-second (Tbps). In terms of the optical communication, THz communication is more favorable to adapt the inconvenient climate conditions including fog, dust, snow and so on. Besides, THz frequency band has the possibility to employ the reflected paths to enhance the communication performance in the complex propagation environments. Stated thus, the THz frequency band is envisioned as the most appropriate candidate to enable the 6G communication systems.

Except for the advantages mentioned above, THz communication still faces some challenges when it is applied to the indoor application scenario. Due to the serious path attenuation and strong directivity of THz waves, the line-of-sight (LOS) THz path can be easily blocked by the indoor obstacles (e.g., walls, doors, wardrobes, human bodies) [7], [8]. Thus, the non-line-of-sight (NLOS) paths play a main role in such a communication scenario. To tackle this issue, an intelligent reflecting surface (IRS) is newly proposed to improve the spectral efficiency and enhance the coverage performance recently [9]–[16]. The IRS is a physical meta-surface consisting of a large number of small-unit reflectors. In general, the IRS is equipped with a number of passive elements and is controlled by a central processor. Each element of the IRS is able to reflect incident THz waves independently with an adjustable phase-shift. Different from the existing candidate techniques, such as conventional reflect-arrays [17], [18], amplify-and-forward relaying [19]–[21] and backscatter communication [22]–[24], IRS owns some distinct advantages. In contrast to reflect-arrays and backscatter communication that require the radio frequency (RF) chains to realize the phase shifts, the hardware complexity of IRS is much lower due to the lack of RF chains, and thus much more reflecting elements can be scaled up in practice. Additionally, since IRS is designed with passive reflecting elements, IRS is also an energy efficient technique. Instead, the amplify-and-forward relaying system usually works with high power consumption. Based on the inherent characteristics, IRS has become a popular research priority extensively.

To the best of our knowledge, the initial concept about IRS is an intelligent wall, which is regarded as an autonomous part of a smart indoor environment to control radio coverage by make use of active frequency-selective surfaces [25]–[27]. In [28], the reconfigurable reflect-array which has the ability to enhance the strength of received signal and eliminate interference simultaneously, is proposed to increase the spectrum sharing capacity for indoor environments. Recently, the IRS-enhanced multiple-input single-output (MISO) wireless systems are explore to maximize the total received power at the user side by jointly optimizing the active beamforming at the base station (BS) and passive beamforming at the IRS [29]–[31]. However, these research contents consider that the phase shifts of the IRS elements are continuous, and the IRS with continuous phase shift results in high fabrication cost and complex hardware design. Later, in [13], [32], the IRS with only a limited number of phase shifts is adopted optimize the energy efficiency and sum-rate performance for both the discrete phase shifts of IRS elements and transmit power allocation. It is worth noting that all the prior work assumes that the channel status information (CSI) is available at the BS side. In other words, we need to solve the channel estimation problem firstly. The main challenge of the channel estimation in IRS-assisted communication system is that the IRS elements are passive and are unable to process the incident signals. Given this, the authors in [33] propose a novel IRS structure that combines the passive elements with a few active elements. These active elements are connected to the baseband processor, and then are used to obtain the whole phase shifts of IRS elements by training a small number of active elements. Except for the low channel estimation accuracy, these active elements increase the power consumption of the IRS-assisted communication system. Instead, the authors present a cascaded channel estimation method with only passive reflecting elements [34]. Nevertheless, this method does not consider the channel characteristics (e.g., sparsity, path loss) in THz band, and it is unable to estimate the BS-IRS channel and IRS-user channel simultaneously. Through the above discussion, the IRS-assisted THz communication system is still worthy of further exploration.

In this paper, we present an IRS-assisted THz multiple-input multiple-output (MIMO) communication system for 6G indoor application scenarios. Considering the hardware design of IRS in THz band, the tunable graphene, which is a two-dimensional carbon material, is selected to achieve a wide phase response range of IRS elements by controlling the applied voltage. With the given hardware structure of graphene-based IRS, the IRS-assisted THz communication system is modeled firstly. Next, before the data transmission, it is necessary to acquire the CSI according to the system model where the geometric channel model is adopted that contains few limited scattering paths [35]–[37]. Then, by exploiting the sparsity feature of the THz channel, the compressed sensing (CS) technique [38] can be utilized to estimate the channel parameters, including path gain, angle of arrival (AoA) and angle of departure (AoD) [39]. With the estimated CSI, we then optimize the communication rate maximization problem by searching the optimal phase shift combination of the IRS elements. Furthermore, three different phase shift search methods are investigated, such as exhaustive search method, local search method and the deep neural network (DNN) based method. Finally, the communication rate performance and the complexity analysis of diverse phase shift search methods are also provided in this paper. The major contributions can be summarized as follows:

- To begin with, to obtain a wide phase response range of reflecting elements, the electric properties of graphene at THz frequency band is investigated by using the CST Microwave Studio which confirms that the tunable graphene is an effective material to design IRS. Based on the electric characteristics of the graphene, the hardware structure of graphene-based IRS is creatively proposed in this paper, where the phase shift of each reflecting element belongs to a discrete phase set. With the designed IRS architecture, the IRS-assisted THz communication model is derived in detail.

- To realize the CSI acquisition of the IRS assisted THz system, an iterative atom pruning based subspace pursuit (IAP-SP) scheme with low complexity is proposed. Compared with the classical subspace pursuit (SP) method [40], the proposed IAP-SP is able to reduce the dimension of the sensing matrix without any recovery performance loss, where the columns (atoms) of the sensing matrix that least correlate with the signal residual tends to be eliminated in the iterative process.

- With the estimated channel information, we note that the communication rate maximization problem of IRS-assisted THz system is a non-convex problem, and then we convert this NP-hard problem to the phase shift search problem. Although the exhaustive search method achieves the optimal performance, it inherits the really high complexity. Fortunately, the local search method is proposed to decrease the computational burden by searching the partial phase shift combinations.

- Moreover, in order to fully explore the relationship between the channel features and the phase shift selection of IRS elements, the DNN-based phase shift search method is employed in this paper to optimize the communication rate. Concretely, the proposed DNN-based search method has the ability to unveil the mapping between the channel features and the phase shift selection. Thus, the DNN-based search method can realize the near-optimal communication rate performance with extremely low complexity.

The rest of this paper is organized as follows. In Section II, we introduce the hardware structure of tunable graphene-based IRS, and the simulation analysis of tunable graphene properties are also presented. In Section III, the IRS-assisted THz communication system and channel model are presented. Also, the channel estimation problem is settled. In Section IV, the communication rate maximization problem is formulated and three different solutions are proposed to search the optimal phase shift combination. Then the simulation results are presented in Section V to verify the proposed algorithms. Finally, we conclude this paper in Section VI.

In this paper, we use the following notation: A is a matrix, a is a vector, a is a scalar. $\|A\|_F$ is the Frobenius norm, whereas A^H, A^*, A^T, A^{-1}, $A^\dagger$ and $rank(A)$ are conjugate transpose, conjugate, transpose, inverse, pseudo-inverse and the rank of A, respectively. $A \otimes B$ is the Kronecker product of A and B. $diag(a)$ is a diagonal matrix with elements of a on its diagonal. $Tr(A)$ is the trace of matrix A. $E[\cdot]$ is used to denote the expectation. $\Re\{\cdot\}$ is the operation to extract the real part of a vector, and $\Im\{\cdot\}$ is the operation to extract the image part of a vector. $supp(a)$ denotes the set of non-zero elements in vector a. $vec(A)$ is the column-ordered vectorization of matrix A, and $vec^{-1}(A)$ is the reverse operation of $vec(A)$. $A[\Omega]$ represents the sub-matrix of A that only contains those columns contained in the index set Ω. $\Omega = supp(a)$ denotes the index set corresponding to non-zero elements of vector a. $\mathcal{M}(a, K)$ is a index set that corresponds to the largest K elements of a. $\mathcal{Z}(a)$ is a index set that corresponds to the zero elements of a, and $\mathcal{Z}^{-1}(a)$ is the reverse operation of $\mathcal{Z}(a)$, respectively.

II. HARDWARE PRELIMINARIES OF IRS

In this section, the design theory and working principle of IRS will be explained, and an achievable structure of IRS is proposed, which is suitable for THz communication systems. From the perspective of hardware structure, IRS employed in THz communication scenarios is a kind of tunable metasurface, which refers to a series of artificial structures with special physical properties, such as negative refractive index, near-zero index and so on [41]. The IRS are composed of periodically arranged IRS elements (also namely meta-atoms), and each element can control the amplitude and phase of the reflected waves. It is worth noting that the size of each IRS element is at a subwavelength level, and thus the propagation state of the reflected waves can be controlled through the arrangement of IRS elements. In general, the structure of each IRS element is designed first to attain the amplitude and phase control of reflected waves, and then these reflecting elements will be arranged properly to construct the reflective wavefront.

A. ELECTRIC PROPERTIES OF GRAPHENE IN THz BAND

As previously discussed, we know that IRS is a kind of tunable metasurface. In order to actively control the reflected waves, the existing devices (e.g., varactor [42], transistor [43]) can be integrated with the structure design. However, due to the short wavelength of the THz waves, the size of each IRS element becomes much smaller and it is difficult to combine those semiconductor devices with IRS elements. In this case, adopting the novel tunable materials, such as graphene, is a feasible choice.

Graphene is a monolayer of carbon atoms arranged in a honeycomb lattice, which has attracted tremendous attention for its extraordinary properties. To verify that the conductivity of graphene can be effectively tuned by applied voltage [44], the complex conductivity of graphene needs to be considered firstly. According to the [45], the complex conductivity of graphene consists of interband and intraband transition contributions, where the interband transition contribution can be neglected and the intraband transition contribution plays a primary role. Then, the intraband transition contribution of

the complex conductivity of graphene can be given as

$$\sigma = \frac{2e^2}{\pi \hbar^2} k_B T \cdot \ln \left[2 \cosh \left(\frac{E_F}{2k_B T} \right) \right] \frac{i}{\omega + i\tau^{-1}}, \quad (1)$$

where ω is the angular frequency, e is the elementary charge, $\hbar$ is the reduced Planck constant, k_B is the Boltzmann constant, T is the temperature, E_F is the Fermi level and $\tau = 1 \times 10^{-12} s$ is the relaxation time, respectively. In terms of the THz frequency band, we know that the conductivity of graphene is mainly dependent on intraband transition, which is related to Fermi energy [46]. Furthermore, the Fermi energy can be formulated as

$$|E_F| = \hbar v_F \left(\frac{\pi \varepsilon V_g}{e t_s} \right)^{\frac{1}{2}}, \quad (2)$$

where the $v_F = 1 \times 10^6 \ m \cdot s^{-1}$ is the Fermi velocity, ε is the dielectric constant of the substrate, t_s refers to its thickness and V_g is the applied voltage, respectively. From the formulation above, it is clear that the conductivity of graphene can be controlled by applying different voltages.
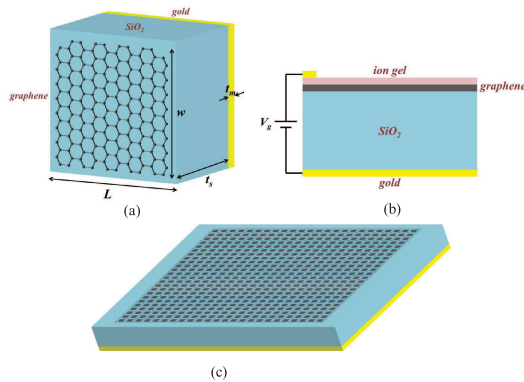

FIGURE 1. The hardware schematic of IRS. (a) The structure of each reflecting element, where $L = 200\mu m$, $w = 190\mu m$, $t_s = 80\mu m$, $t_m = 0.2\mu m$. The permittivity of SiO_2 is 3.75; (b) The theoretical principle of applying voltage to graphene; (c) An indicative array of IRS elements.

B. HARDWARE DESIGN OF GRAPHENE-BASED IRS

Fig. 1 illustrates a typical hardware structure of graphene-based IRS consisting of a array of reflecting elements. These subwavelength elements are closely arranged and only a single element is unable to realize the function of phase control due to the scattering of THz waves. For example, the size of IRS is $20mm \times 20mm$ while such an array contains 100×100 reflecting elements. From Fig. 1 (a) we can see that, each reflecting element of IRS comprises the graphene patch, the silicon dioxide substrate and the gold ground plane. This hardware structure of an IRS can be considered as a Fabry-Perot resonance cavity [47], and the reflecting phase can be written as

$$\varphi_r = m\pi - wk_0 Re\left(n_{sp}\right), \quad (3)$$

where w is the width of graphene patch, $k_0 = 2\pi / \lambda_0$ is the wave number of free space, φ is the phase shift of reflection wave, m is an integer, n_{sp} is the effective refraction index of the surface plasmon, which is the function of graphene's effective permittivity ε_{eff} and the angular frequency ω. The permittivity of the graphene can be calculated as

$$\varepsilon_{eff} = 1 + \frac{i\sigma}{\omega \varepsilon_0 t_g}, \quad (4)$$

where t_g is the thickness of graphene. According to (3) and (4), we can obtain the relationship between the phase response and graphene's conductivity, that is, the phase shift of each IRS element can be altered by changing graphene's conductivity through applied voltage. Fig. 1 (b) shows a schematic diagram of controlling each reflecting element. The attached material, termed ion-gel, is used to connect graphene and the electrode, and the voltage V_g is applied between the electrode and the ground plane.
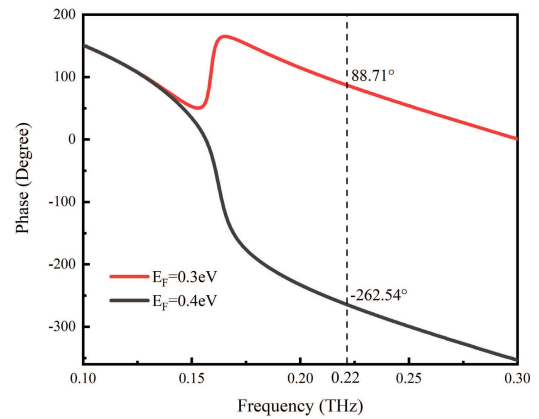

FIGURE 2. Simulation results of phase response with $E_F = 0.3$ eV and 0.4 eV when the incident wave is at 0.22 THz.

To verify the effectiveness of the graphene-based hardware structure, simulation results of each IRS element by using CST Microwave Studio is shown in Fig. 2. Fig. 2 depicts the phase response of the IRS elements along with the frequency and the diverse Fermi level values. One may note that the proposed IRS just work well at a narrow frequency band (about 3 GHz) once the hardware structure of IRS is fabricated. For a normally incident wave at 0.22 THz, the controllable phase shift range of each IRS element is able to reach the phase range of 351.25° with $E_F = 0.3$ eV producing a phase of 88.71° and 0.4 eV generating a phase of −262.54°. More importantly, by properly adjusting the parameters of hardware structure, like the length of the substrate L and the width of the graphene patch w, the available phase shift range can be consistent with the interval $[0, 2\pi]$ [48]. Remarkably, we can take any desired phase shift from the phase range by applying the voltages continuously, which results in high complexity hardware circuits and expensive system overhead in practice.

Based on the above analysis about the hardware characteristics of IRS, we indicate that the phase-shift φ of each IRS element can be adjusted within the range $[\varphi_{\min}, \varphi_{\max}]$,
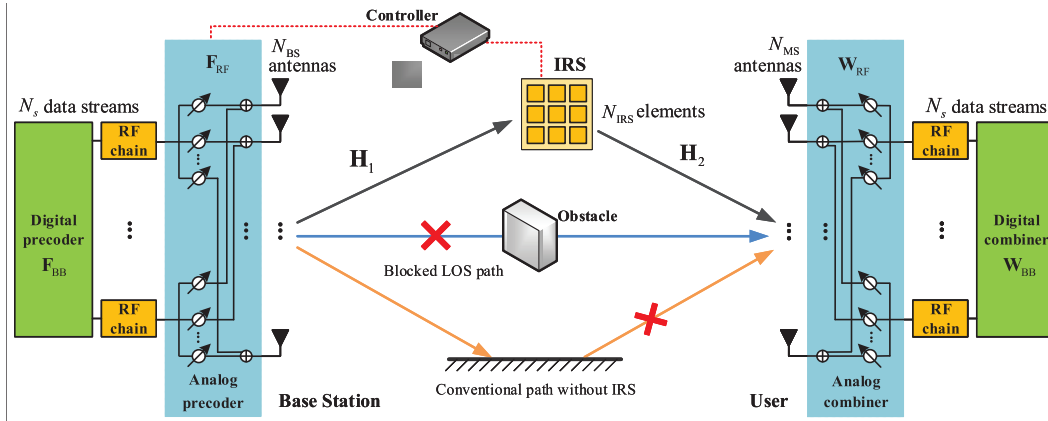
FIGURE 3. IRS-assisted THz MIMO communication system.

where $\varphi_{\min}$ is the controllable minimum phase shift and $\varphi_{\max}$ is the controllable maximum phase shift for a given IRS structure. Since the power consumption and hardware cost of IRS with discrete phase-shifts is much more practical than the IRS with continuous phase-shifts, we therefore consider discrete structure of the phase-shifts in our work. The phase shift of each reflecting element φ belongs to a finite phase set $\mathcal{F} = \{\varphi_{\min}, \varphi_{\min} + \Delta\varphi \cdots, \varphi_{\min} + \Delta\varphi \left(2^b - 1\right)\}$, where b is a bit quantization number and $\Delta\varphi = (\varphi_{\max} - \varphi_{\min})/2^b$ is the phase interval. For example, when $b = 2$ bits, $\varphi_{\min} = 0$, and $\varphi_{\max} = 2\pi$, $\mathcal{F} = \{0, \pi/2, \pi, (3\pi)/2\}$.

III. SYSTEM MODEL AND CHANNEL ESTIMATION

In this section, the IRS-assisted THz communication model is introduced from a MIMO perspective. Then, the channel estimation problem and the corresponding solution are formulated in the following.

A. SYSTEM MODEL

Consider a downlink MIMO THz communication system and the communication link is assisted with IRS that consists of N_{IRS} passive reflecting elements, as shown in Fig. 3. The hybrid beamforming architecture is adopted, since we take the hardware implementation and system cost into consideration. Assume that a BS with N_{BS} antennas and M_{BS} RF chains serves a single mobile station (MS) with N_{MS} antennas and M_{MS} RF chains. In general, the number of antennas is larger than the number of the RF chains due to the serious power consumption of the RF circuits, i.e. $N_{BS} > M_{BS}$, $N_{MS} > M_{MS}$. There are N_s data streams that are transmitted between BS and MS. Due to the complex environment of indoor communication scenarios, the LOS path between BS and MS is usually blocked by the obstacles and NLOS paths play a major role. Also, the NLOS paths that are just reflected by an IRS for the first time are considered here. Furthermore, a time-division duplexing (TDD) scheme can be used for the downlink CSI acquisition by utilizing the characteristic of channel reciprocity.

For the channel training scheme in such a downlink MIMO system, we suppose that the BS uses P different precoding vectors at P successive time frame. Each time frame contains Q time slots, and a combining vector is employed by the MS to detect the transmitted signal at each time slot. At pth time frame, the transmitted beamforming vector at the BS side can be expressed as

$$x_p = F_p s_p = F_p^{RF} F_p^{BB} s_p, \tag{5}$$

where $s_p \in \mathcal{C}^{N_s}$ is the transmitted pilot symbol vector and F_p is the $N_{BS} \times N_s$ BS precoding matrix that combines the RF precoding vector $F_p^{RF} \in \mathcal{C}^{N_{BS} \times M_{BS}}$ and the baseband precoding vector $F_p^{BB} \in \mathcal{C}^{M_{BS} \times N_s}$. Note that the pilot vector satisfies $E[s_p s_p^H] = \left(P_s/N_s\right) I_{N_s}$ and P_s is the total transmitted power of N_s pilot symbols.

At the receiver side, the MS successively adopts Q combining vectors to detect the transmitted signal. For a given time frame p, the MS employs a combining vector w_q to combine the received signal at time slot q, then the resulting signal formulation is written as

$$y_{q,p} = w_q^H H x_p + z_{q,p}, \tag{6}$$

where $w_q \in \mathcal{C}^{N_{MS}}$ is the combining vector at qth time slot, $z_{q,p}$ is the additive Gaussian noise and $H \in \mathcal{C}^{N_{MS} \times N_{BS}}$ is the cascaded channel matrix, respectively. More importantly, the cascaded channel H is composed of the BS-IRS channel $H_1 \in \mathcal{C}^{N_{IRS} \times N_{BS}}$ and the IRS-MS channel $H_2 \in \mathcal{C}^{N_{MS} \times N_{IRS}}$, which differs from conventional channel model without IRS and will be later introduced in this paper. After collecting the received signals processed by Q combining vectors, the received signal vector at pth time frame can be formulated as

$$y_p = W^H H x_p + z_p = W^H H F_p^{RF} F_p^{BB} s_p + z_p, \tag{7}$$

where

$$\begin{aligned} y_p &= [y_{1,p}, y_{2,p}, \cdots, y_{Q,p}]^T \\ z_p &= [z_{1,p}, z_{2,p}, \cdots, z_{Q,p}]^T \\ W &= [w_1, w_2, \cdots, w_Q] \end{aligned} \tag{8}$$

With regard to the total P time frames, the MS utilizes the identical combiner $\boldsymbol{W} \in \mathcal{C}^{N_{MS} \times Q}$ to detect the received signals, and thus, the received matrix at the MS side can be written as

$$Y = W^H H X + Z, \qquad (9)$$

where $\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_P]$ is a $Q \times P$ received matrix, $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_P]$ is a $N_{BS} \times P$ hybrid beamforming matrix, $\boldsymbol{Z} = [\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_P]$ is a $Q \times P$ noise matrix by concatenating the P noise vectors, respectively.

B. CHANNEL MODEL

Before formulating the channel estimation problem, the channel model for the IRS-assisted MIMO THz communication system should be described firstly. Different from the conventional channel model [39], [49], [50], the channel model of the IRS-based THz system includes the BS-IRS channel $\boldsymbol{H}_1$, the IRS-MS channel $\boldsymbol{H}_2$, and the phase shift matrix $\boldsymbol{\Phi} = diag\left(\left[e^{j\varphi_1}, \cdots, e^{j\varphi_{N_{IRS}}}\right]^T\right)$, where $\boldsymbol{\Phi}$ is a $N_{IRS} \times N_{IRS}$ diagonal matrix to indicate the phase shifts of IRS elements. Then, the cascaded channel of the IRS assisted system can be expressed as

$$H = H_2 \Phi H_1. \qquad (10)$$

Without loss of generality, in THz frequency band, we assume that both $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ are able to be presented by the geometric channel model with few scattering paths. Furthermore, each scattering path is supposed to decide a single propagation path. Based on the geometric channel model, the BS-IRS channel $\boldsymbol{H}_1$ can be formulated as

$$H_1 = \sqrt{\frac{N_{BS} N_{IRS}}{L_1}} \sum_{l_1=1}^{L_1} \alpha_{l_1} \boldsymbol{a}_{IRS}(\gamma_{l_1}^{AOA}) \boldsymbol{a}_{BS}^H(\phi_{l_1}), \qquad (11)$$

where L_1 denotes the scattering paths of BS-IRS link, α_{l_1} is the complex gain of l_1th path, $\gamma_{l_1}^{AOA} \in [0, 2\pi]$ is the AoA and $\phi_{l_1} \in [0, 2\pi]$ is the AoD of l_1th path corresponding to BS and IRS, respectively. The complex gains of the propagation paths are mainly influenced by the path loss and molecular absorption, and more details are recommended in [51]. For simplification, the uniform linear arrays (ULAs) are considered here. Hence, we can get the array response vectors $\boldsymbol{a}_{BS}(\phi_{l_1}) \in \mathcal{C}^{N_{BS}}$ and $\boldsymbol{a}_{IRS}(\gamma_{l_1}^{AOA}) \in \mathcal{C}^{N_{IRS}}$ at the BS and IRS. Specifically, the array response vectors $\boldsymbol{a}_{BS}(\phi_{l_1})$ and $\boldsymbol{a}_{IRS}(\gamma_{l_1}^{AOA})$ can be written as follow

$$\boldsymbol{a}_{BS}(\phi_{l_1}) = \frac{1}{\sqrt{N_{BS}}} \left[1, e^{j(2\pi/\lambda)d \sin(\phi_{l_1})}, \cdots,\right.$$
$$\left. e^{j(N_{BS}-1)(2\pi/\lambda)d \sin(\phi_{l_1})}\right]^T, \qquad (12)$$

$$\boldsymbol{a}_{IRS}(\gamma_{l_1}^{AOA}) = \frac{1}{\sqrt{N_{IRS}}} \left[1, e^{j(2\pi/\lambda)d \sin(\gamma_{l_1}^{AOA})}, \cdots,\right.$$
$$\left. e^{j(N_{IRS}-1)(2\pi/\lambda)d \sin(\gamma_{l_1}^{AOA})}\right]^T, \qquad (13)$$

where λ is the wavelength of the THz signals and d is the distance between adjacent antenna elements or IRS elements, which is usually defined as $d = \lambda/2$. Apart from the normal expression (11), a more compact form of BS-IRS channel can be formulated as

$$H_1 = A_{IRS,1} diag(\boldsymbol{\alpha}) A_{BS}^H, \qquad (14)$$

where $\boldsymbol{\alpha} = \sqrt{N_{BS} N_{IRS}/L_1} \left[\alpha_1, \alpha_2, \cdots, \alpha_{L_1}\right]^T$, and the array response matrices can be respectively given as

$$A_{BS} = \left[\boldsymbol{a}_{BS}(\phi_1), \cdots, \boldsymbol{a}_{BS}(\phi_{L_1})\right], \qquad (15)$$

$$A_{IRS,1} = \left[\boldsymbol{a}_{IRS}(\gamma_1^{AOA}), \cdots, \boldsymbol{a}_{IRS}(\gamma_{L_1}^{AOA})\right]. \qquad (16)$$

Similar to the BS-IRS channel $\boldsymbol{H}_1$, the IRS-MS channel $\boldsymbol{H}_2$ can be formulated as

$$H_2 = A_{MS} diag(\boldsymbol{\beta}) A_{IRS,2}^H, \qquad (17)$$

where

$$\boldsymbol{\beta} = \sqrt{N_{MS} N_{IRS}/L_2} \left[\beta_1, \beta_2, \cdots, \beta_{L_2}\right]^T,$$
$$A_{MS} = \left[\boldsymbol{a}_{MS}(\theta_1), \cdots, \boldsymbol{a}_{MS}(\theta_{L_2})\right],$$
$$A_{IRS,2} = \left[\boldsymbol{a}_{IRS}(\gamma_1^{AOD}), \cdots, \boldsymbol{a}_{IRS}(\gamma_{L_2}^{AOD})\right]. \qquad (18)$$

By substituting (14) and (17) into (10), the entire channel model can be written as

$$H = A_{MS} diag(\boldsymbol{\beta}) A_{IRS,2}^H \Phi A_{IRS,1} diag(\boldsymbol{\alpha}) A_{BS}^H. \qquad (19)$$

C. PROPOSED CHANNEL ESTIMATION SCHEME

In this subsection, the channel estimation problem and the corresponding solution of IRS-aided THz MIMO system is to introduced in detail. The main challenge is to estimate the BS-IRS channel and the IRS-MS channel simultaneously. With the geometric channel model mentioned above, we discover that the channel estimation problem can be converted to the sparse recovery problem by utilizing the poor scattering characteristics of THz channel. Then, the CS technique is an efficient tool to handle such a sparse problem.

To be specific, before formulating the spare problem, we need to vectorize the received matrix $\boldsymbol{Y}$ in (9), which can be expressed as

$$\boldsymbol{y}_v = vec(\boldsymbol{Y})$$
$$= \left(\boldsymbol{X}^T \otimes \boldsymbol{W}^H\right) vec(\boldsymbol{H}) + vec(\boldsymbol{Z}). \qquad (20)$$

For the sake of forming the CS problem, the grid quantization procedure is achieved necessarily [39]. Here, we consider that the AoAs and AoDs of $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ are selected from a uniform grid of N_1 ($N_1 >> L_1$) points and a uniform grid of N_2 ($N_2 >> L_2$) points, respectively. Once the grid quantization procedure of azimuth angles is completed, the quantized AoAs and AoDs of $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ can be described, i.e., $\bar{\phi}_{l_1}$, $\bar{\gamma}_{l_1}^{AOA} = 2\pi l_1/L_1$, $l_1 \in [0, L_1 - 1]$, $\bar{\theta}_1$, $\bar{\gamma}_1^{AOD} = 2\pi l_2/L_2$, $l_2 \in [0, L_2 - 1]$. According to (15)

and (16), the quantized array response matrices can be defined as

$$
\begin{aligned}
\bar{A}_{BS} &= \left[a_{BS}(\bar{\phi}_1), \cdots, a_{BS}(\bar{\phi}_{N_1}) \right], \\
\bar{A}_{MS} &= \left[a_{MS}(\bar{\theta}_1), \cdots, a_{MS}(\bar{\theta}_{N_2}) \right], \\
\bar{A}_{IRS,1} &= \left[a_{IRS}(\bar{\gamma}_1^{AOA}), \cdots, a_{IRS}(\bar{\gamma}_{N_1}^{AOA}) \right], \\
\bar{A}_{IRS,2} &= \left[a_{IRS}(\bar{\gamma}_1^{AOD}), \cdots, a_{IRS}(\bar{\gamma}_{N_2}^{AOD}) \right].
\end{aligned}
\tag{21}
$$

Subsequently, the cascaded channel model (19) after the grid quantization operation can be rewritten as

$$
H = \bar{A}_{MS} G_2 \bar{A}_{IRS,2}^H \Phi \bar{A}_{IRS,1} G_1 \bar{A}_{BS}^H,
\tag{22}
$$

where $G_1 \in \mathcal{C}^{N_1 \times N_1}$ is the quantized path gain matrix of H_1 and $G_2 \in \mathcal{C}^{N_2 \times N_2}$ is the quantized path gain matrix of H_2, respectively. It is worth noting that the non-zero elements of G_1 are equivalent to the elements of α due to the grid quantization principle [52]. Also, such a mapping relationship is available for G_2 and β.

Combining (20) and (22), the sparse formulation of CS problem can be expressed as

$$
\begin{aligned}
y_v &= \left(X^T \otimes W^H \right) \Big[\left((vec\,(\Phi))^T \left(\bar{A}_{IRS,1} \otimes \bar{A}_{IRS,2}^* \right) \right) \\
&\quad \otimes \left(\bar{A}_{BS}^* \otimes \bar{A}_{MS} \right) \Big] vec \left(G_1^T \otimes G_2 \right) + vec\,(Z) \\
&= A_v g_v + z_v,
\end{aligned}
\tag{23}
$$

where $A_v = \left(X^T \otimes W^H \right) \Big[\left((vec\,(\Phi))^T \left(\bar{A}_{IRS,1} \otimes \bar{A}_{IRS,2}^* \right) \right)$ $\otimes \left(\bar{A}_{BS}^* \otimes \bar{A}_{MS} \right) \Big]$ is observed as $PQ \times N_1^2 N_2^2$ sensing matrix, $g_v = vec \left(G_1^T \otimes G_2 \right)$ is a $N_1^2 N_2^2 \times 1$ vector with $L_1 L_2$-sparse and $z_v = vec\,(Z)$ is a $PQ \times 1$ vectorized noise vector, respectively. Based on structure properties of (23), it can be treated as a sparse recovery problem, and thus different CS techniques are able to be leveraged to estimate the channel parameters, such as SP, matching pursuit (MP) [53], orthogonal matching pursuit (OMP) [54], [55], and iterative hard thresholding (IHT) [56].

Proof: The detailed proof for the above (23) is provided in Appendix A. ∎

In terms of the sparse reconstruction problem (23), the CS technique involves the unacceptably computational complexity due to the high dimension of sensing matrix A_v. For instance, we suppose $P = 6, Q = 12, N_1 = 30, N_2 = 30$, then the dimension of A_v is 72×30^4. It is worth noting that the high-dimensional sensing matrix is caused by the double grid quantization operation, where the double grid quantization means that the grid quantization of H_1 and the grid quantization of H_2 are operated synchronously. Thus, the critical objective is to reduce the dimension of A_v as well as the complexity.

In this paper, the SP algorithm is employed to perform the better recovery performance, which is able to both remove and add the elements in an active set. To efficiently strike the computational difficulty, we propose a low complexity IAP-SP algorithm. In recovery process of classical SP,

we find that plenty of columns of the sensing matrix A_v have no correlation with the signal residual. In other words, partial elements of the resulting vector that combines the sensing matrix and the residual are zero. Consequently, during iterative process, it is redundant to make the columns of A_v with no contributions to the signal residual participate in calculations. According to this phenomenon, the proposed IAP-SP algorithm attains an attractive complexity reduction by iteratively eliminating the redundant columns in sensing matrix. The proposed low complexity IAP-SP algorithm is described in Algorithm 1.

Algorithm 1 Proposed IAP-SP Algorithm

Require: Vectorized received signal y_v, sensing matrix A_v, sparsity $K = L_1 L_2$, number of time frames P, number of time slots Q.

1: Initialization: the support set $\Gamma^{(t-1)} = \emptyset, A_v^{(0)} = A_v$, iteration index $t = 1$, the residual vector $r^{(t-1)} = y_v$, the redundant column index set of A_v: $\Omega^{(t-1)} = \emptyset$, $\Xi = \left\{ 1, 2, \cdots, N_1^2 N_2^2 \right\}$.

   Iterative Process:

2: Support set estimation:
   $\bar{\Gamma}^{(t)} = \Gamma^{(t-1)} \cup \mathcal{M} \left(\left(A_v \left[\Xi \backslash \Omega^{(t-1)} \right] \right)^H r^{(t-1)}, K \right)$,

3: Redundant column index set update:
   $\Omega^{(t)} = \Omega^{(t-1)} \cup \mathcal{Z} \left(\left(A_v \left[\Xi \backslash \Omega^{(t-1)} \right] \right)^H r^{(t-1)} \right)$,

4: Least square estimation:
   $u \left[\bar{\Gamma}^{(t)} \right] = \left(A_v \left[\bar{\Gamma}^{(t)} \right] \right)^{\dagger} y_v$,

5: Support pruning:
   $\Gamma^{(t)} = \mathcal{M} \left(u \left[\bar{\Gamma}^{(t)} \right], K \right)$,

6: Estimated recovery vector:
   $\tilde{g}_v^{(t)} \left[\Gamma^{(t)} \right] = \left(A_v \left[\Gamma^{(t)} \right] \right)^{\dagger} y_v, \tilde{g}_v \left[\Xi \backslash \Gamma^{(t)} \right] = 0$,

7: Residual update:
   $r^{(t)} = y_v - \left(A_v \left[\Gamma^{(t)} \right] \right) \tilde{g}_v \left[\Gamma^{(t)} \right]$,

8: **if** $\left\| r^{(t)} \right\| < \left\| r^{(t-1)} \right\|$, **then**
   $t = t + 1$,
   return to step 2,
   end if

9: Compressed recovery vector:
   $\bar{g}_v^{(t)} = \left(A_v^{(t-1)} \left[\Gamma^{(t-1)} \right] \right)^{\dagger} y_v$,

10: Output estimated vector:
   $\bar{g}_v = \mathcal{Z}^{-1} \left(\tilde{g}_v^{(t-1)} \right)$ and satisfying $\bar{g}_v \left[\Xi \backslash \Gamma^{(t-1)} \right] = 0$,

Ensure: Recovered signal vector $\bar{g}_v = vec \left(\bar{G}_1^T \otimes \bar{G}_2 \right)$.

Now, some important steps of Algorithm 1 are to be explained in the following. Firstly, as for step 1-2, the algorithm initialization and the support set estimation are presented, respectively. Secondly, in step 3, the index set of the redundant columns in A_v is updated, where these redundant columns are not involved in subsequent calculations. It is worth mentioning that step 3 is a very pivotal operation in our proposed low complexity IAP-SP algorithm. Then, for steps 4-9, the operation principle of the proposed low complexity IAP-SP scheme is the same as classical SP algorithm

in order to find the K-sparse solution to (23). Finally, for step 10, we need to fill the zero-elements to the corresponding positions that is caused by step 3. According to the analysis of the algorithm steps, the computational complexity of classical SP algorithm and our proposed IAP-SP in terms of complex multiplications can be expressed as

$$\mathcal{O}_{SP} \begin{pmatrix} T\left(N_1^2 N_2^2 + 3L_1 L_2\right) PQ + 2TP^3 Q^3 \\ +4TL_1 L_2 P^2 Q^2 \end{pmatrix} \quad (24)$$

$$\mathcal{O}_{IAP-SP} \begin{pmatrix} \left(\sum_{t=1}^{T} \left|\Omega^{(t)}\right|\right) PQ + 3TL_1 L_2 PQ \\ +2TP^3 Q^3 + 4TL_1 L_2 P^2 Q^2 \end{pmatrix} \quad (25)$$

respectively, where T represents the assumed maximum iterations.

Although the sparse vector $\bar{\boldsymbol{g}}_v$ is reconstructed, estimating the $\bar{\boldsymbol{G}}_1$ and $\bar{\boldsymbol{G}}_2$ separately is still a tricky problem, where the recovered $\bar{\boldsymbol{g}}_v$, $\bar{\boldsymbol{G}}_1$, and $\bar{\boldsymbol{G}}_2$ correspond to actual $\boldsymbol{g}_v$, $\boldsymbol{G}_1$, and $\boldsymbol{G}_2$, respectively. Considering the practical communication scenarios, the positions of BS and IRS are fixed once the communication environment is determined. Therefore, we assume that the sum of the total path gains between BS and IRS is a constant C_0 for a specific environment, where C_0 is decided by the transmit power at BS side. Then we can get the following expression as

$$\sum_{l_1=1}^{L_1} \bar{\alpha}_{l_1} = C_0, \quad (26)$$

where $\bar{\alpha}_{l_1} \in supp\left(vec\left(\bar{\boldsymbol{G}}_1\right)\right)$ and $l_1 \in [1, \cdots, L_1]$. Here, we define $\bar{\boldsymbol{G}}_v = vec^{-1}\left(\bar{\boldsymbol{g}}_v\right)$ from a block-matrix perspective, and then $\bar{\boldsymbol{G}}_v$ can be rewritten as

$$\bar{\boldsymbol{G}}_v = vec^{-1}\left(\bar{\boldsymbol{g}}_v\right) = \bar{\boldsymbol{G}}_1^T \otimes \bar{\boldsymbol{G}}_2$$
$$= \begin{bmatrix} \boldsymbol{A}_{1,1} & \cdots & \boldsymbol{A}_{1,N_1} \\ \vdots & \ddots & \vdots \\ \boldsymbol{A}_{N_1,1} & \cdots & \boldsymbol{A}_{N_1,N_1} \end{bmatrix} \quad (27)$$

where $\boldsymbol{A}_{i_1,j_1} \in \mathcal{C}^{N_2 \times N_2}$, $\forall i_1, j_1 \in [1, N_1]$.

Combining (26) and (27), the matrix $\bar{\boldsymbol{G}}_2$ can be derived as

$$\bar{\boldsymbol{G}}_2 == \frac{\sum_{i_1=1}^{N_1} \sum_{j_1=1}^{N_1} \boldsymbol{A}_{i_1,j_1}}{C_0}, \quad (28)$$

Suppose $a_{i_1,j_1}(i_2,j_2) \in \boldsymbol{A}_{i_1,j_1}$ and $\bar{g}_{i_2,j_2}^2 \in \bar{\boldsymbol{G}}_2$, where $\forall i_2, j_2 \in [1, N_2]$. Subsequently, the desired matrix $\bar{\boldsymbol{G}}_1$ can be formulated as

$$\bar{\boldsymbol{G}}_1 = \begin{bmatrix} \dfrac{\sum_{i_2=1}^{N_2}\sum_{j_2=1}^{N_2} a_{1,1}(i_2,j_2)}{\sum_{i_2=1}^{N_2}\sum_{j_2=1}^{N_2} \bar{g}_{i_2,j_2}^2} & \cdots & \dfrac{\sum_{i_2=1}^{N_2}\sum_{j_2=1}^{N_2} a_{1,N_1}(i_2,j_2)}{\sum_{i_2=1}^{N_2}\sum_{j_2=1}^{N_2} \bar{g}_{i_2,j_2}^2} \\ \vdots & \ddots & \vdots \\ \dfrac{\sum_{i_2=1}^{N_2}\sum_{j_2=1}^{N_2} a_{N_1,1}(i_2,j_2)}{\sum_{i_2=1}^{N_2}\sum_{j_2=1}^{N_2} \bar{g}_{i_2,j_2}^2} & \cdots & \dfrac{\sum_{i_2=1}^{N_2}\sum_{j_2=1}^{N_2} a_{N_1,N_1}(i_2,j_2)}{\sum_{i_2=1}^{N_2}\sum_{j_2=1}^{N_2} \bar{g}_{i_2,j_2}^2} \end{bmatrix}^T \quad (29)$$

Proof: The detailed proofs for (28) and (29) are provided in Appendix B and C, respectively. ∎

Once the path gain matrices $\bar{\boldsymbol{G}}_1$ and $\bar{\boldsymbol{G}}_2$ are completed, the AoAs and AoDs can be determined as well as the BS-IRS channel $\bar{\boldsymbol{H}}_1$ and IRS-MS channel $\bar{\boldsymbol{H}}_2$ by utilizing the regulation of grid quantization accordingly. For evaluating the channel estimation performance, the normalized mean squared error (NMSE) is defined to compare the validity of diverse sparse recovery methods, which can be formulated as

$$NMSE = \frac{\left\|\boldsymbol{H} - \bar{\boldsymbol{H}}\right\|_F^2}{\left\|\boldsymbol{H}\right\|_F^2}, \quad (30)$$

where $\bar{\boldsymbol{H}}$ is the estimated channel and $\bar{\boldsymbol{H}} = \bar{\boldsymbol{H}}_2 \boldsymbol{\Phi} \bar{\boldsymbol{H}}_1$.
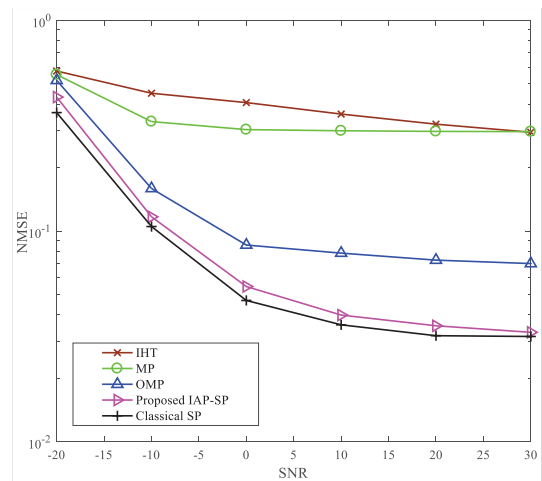
FIGURE 4. NMSE comparison of the considered methods in the IRS-assisted THz communication scenario.

Fig. 4 depicts the NMSE performance comparison among different CS techniques under the main parameter settings of $N_{BS} = 32$, $N_{MS} = 32$, $N_{IRS} = 16$, $P = 6$, and $Q = 12$. Taking the computational burden into consideration, the number of grid points of the grid quantization process is set as $N_1 = N_2 = 30$. From Fig. 4 we note that the classical SP achieves the best sparse reconstruction performance compared to the other considered algorithms, in which the selected atoms are possible to be removed or added to form the estimated support set during the iterative reconstruction process, just as mentioned in [40]. Interestingly, our proposed IAP-SP algorithm is able to realize approximately the same NMSE performance as classical SP, while it outstrips the rest CS techniques, i.e., MP [53], OMP [54] and IHT [56]. Especially, when $SNR = 20$, the NMSE values of the IAP-SP (basically same as classical SP), OMP, MP and IHT are about 3×10^{-2}, 7×10^{-2}, 2.9×10^{-1} and 3.2×10^{-1}, respectively. Additionally, by substituting these parameter settings of the THz communication system into (24) and (25), the proposed IAP-SP algorithm can achieve around 88.79% complexity reduction when the number of iterations is $T = 9$. Through the analysis of computational complexity and recovery performance, the proposed IAP-SP scheme can make a better trade-off between computational

calculations and recovery accuracy compared with classical SP scheme. With the number of the grid quantization points increasing, more complexity reduction is to be attained correspondingly. In addition, according to [57], further improved performance can be achieved by utilizing the redundant dictionary. Therefore, the proposed IAP-SP algorithm is an efficient tool to solve the channel estimation problem of the IRS-assisted THz MIMO communication system.

IV. DATA-RATE MAXIMIZATION FOR IRS-ASSISTED THz COMMUNICATION SYSTEM

With the estimated CSI analyzed above, the major objective is to handle the data-rate maximization problem of the IRS-assisted THz MIMO communication system by jointly designing the IRS matrix $\mathbf{\Phi}$ and the hybrid combiner matrix $\mathbf{W} = \mathbf{W}_{RF}\mathbf{W}_{BB}$, which is an intractable non-convex optimization problem. Given this, we reconsider this NP-hard problem as the phase shift search problem of IRS elements. In terms of the phase shift search issue, we propose three different methods are proposed to optimize the IRS phase-shift matrix, including the exhaustive search scheme, local search scheme and deep learning based scheme. After that, we leverage the classical singular value decomposition (SVD) operation to design the optimal combining matrix $\mathbf{W}$ as well as the maximum communication rate.

Firstly, we present the achievable communication rate of the IRS-aided THz MIMO system as

$$R = \log_2 \left| \mathbf{I}_{N_s} + \frac{\rho}{\delta^2 N_s} \left(\mathbf{W}^H \mathbf{W}\right)^{-1} \mathbf{W}^H \right.$$
$$\left. \times \left(\bar{\mathbf{H}}_2 \mathbf{\Phi} \bar{\mathbf{H}}_1\right) \mathbf{F}\mathbf{F}^H \left(\bar{\mathbf{H}}_2 \mathbf{\Phi} \bar{\mathbf{H}}_1\right)^H \mathbf{W} \right|, \quad (31)$$

where ρ is the total transmission power, δ^2 is the noise power, and $\mathbf{F} = \mathbf{F}_{RF}\mathbf{F}_{BB}$ is the hybrid precoding matrix, respectively. We note that the BS-IRS channel $\bar{\mathbf{H}}_1$ has been estimated in previous section. Considering the practical communication scenarios, the locations of the BS and the IRS are fixed and $\bar{\mathbf{H}}_1$ also maintains unchanged. Given this, the precoder $\mathbf{F}$ at the BS side is a constant matrix and can be obtained by the SVD operation of $\bar{\mathbf{H}}_1$. Specifically, since $\bar{\mathbf{H}}_1$ is able to be decomposed into $\bar{\mathbf{H}}_1 = \mathbf{U}_1 \mathbf{\Lambda}_1 \mathbf{V}_1^H$, then we have the relation as $\mathbf{F} = \tilde{\mathbf{V}}_1$, where $\tilde{\mathbf{V}}_1$ is a sub-matrix with the first N_s column vectors of $\mathbf{V}_1$ [49]. Once the optimal hybrid precoder $\mathbf{F}$ or combiner $\mathbf{W}$ is obtained, the OMP-based solution is adopted to decompose $\mathbf{F}$ or $\mathbf{W}$ into RF part and baseband (BB) part separately [58]. For simplicity, we assume that the IRS satisfies a perfect reflecting mode where each IRS element is set as $\{\mu_i\}_{i=1}^{N_{IRS}} = 1$. Then, the data-rate maximization problem can be written as

$$\left(\mathbf{\Phi}^{opt}, \mathbf{W}^{opt}\right) = \arg \max R,$$
$$s.t. \; \varphi_n \in \mathcal{F}, \quad \forall n = 1, \dots, N_{IRS},$$
$$\mathbf{\Phi} = diag\left(\left[e^{j\varphi_1}, e^{j\varphi_2}, \dots, e^{j\varphi_{N_{IRS}}}\right]^T\right). \quad (32)$$

Based on the above discussion, the optimization objective (32) is a a non-convex and discrete optimization problem with two matrix variables, $\mathbf{\Phi}$ and $\mathbf{W}$. For the sake of solving the

optimization problem (32), existing optimization techniques are hard to operate directly. Nevertheless, it is worth noting that there are still some useful properties existing in problem (32). Firstly, the number of possible candidates of the phase shift matrix $\mathbf{\Phi}$ is finite, because each entry $\{\varphi_n\}_{n=1}^{N_{IRS}}$ is discrete according to the hardware structure of IRS. Secondly, the combining matrix $\mathbf{W}$ is an unconstrained combiner matrix. Moreover, the phase shift matrix $\mathbf{\Phi}$ and the combining matrix $\mathbf{W}$ are mutually independent in the IRS assisted THz communication system.

Motivating by these distinguishing features, one available way of settling the above optimization problem is to first optimize the phase shift matrix $\mathbf{\Phi}$ and later design the combining matrix $\mathbf{W}$. To achieve this, we suppose that the effective channel $\bar{\mathbf{H}}_e$ is defined as $\bar{\mathbf{H}}_e = \bar{\mathbf{H}}_2 \mathbf{\Phi} \bar{\mathbf{H}}_1$. By utilizing the SVD operation, the channel $\bar{\mathbf{H}}_e$ can be decompose into $\bar{\mathbf{H}}_e = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H$, where $\mathbf{U}$ is a $N_{MS} \times rank\left(\bar{\mathbf{H}}_e\right)$ unitary matrix, $\mathbf{\Lambda}$ is a $rank\left(\bar{\mathbf{H}}_e\right) \times rank\left(\bar{\mathbf{H}}_e\right)$ diagonal matrix of singular values arranged in a decreasing order, and $\mathbf{V}$ is a $N_{BS} \times rank\left(\bar{\mathbf{H}}_e\right)$ unitary matrix, respectively. Furthermore, we also define the unitary matrix $\mathbf{U}$ as $\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2]$, and then the optimal combiner is simply given by $\mathbf{W}^{opt} = \mathbf{U}_1$ [49], where $\mathbf{U}_1$ is of dimension $N_{MS} \times N_s$. To realize the better communication performance, we calculate the combiner matrix $\mathbf{W}$ by using the SVD operation of the cascaded channel $\bar{\mathbf{H}}_e$ rather than the IRS-MS channel $\bar{\mathbf{H}}_2$, since the cascaded channel $\bar{\mathbf{H}}_e$ contains the CSI of joint $\mathbf{W}$ and $\mathbf{\Phi}$. Thus, the optimization problem (32) can be rewritten as

$$R = \log_2 \left| \mathbf{I}_{N_s} + \frac{\rho}{\delta^2 N_s} \left(\mathbf{W}^H \mathbf{W}\right)^{-1} \mathbf{W}^H \right.$$
$$\left. \times \bar{\mathbf{H}}_e \mathbf{F}\mathbf{F}^H \bar{\mathbf{H}}_e^H \mathbf{W} \right|,$$
$$s.t. \; \varphi_n \in \mathcal{F}, \quad \forall n = 1, \dots, N_{IRS},$$
$$\mathbf{\Phi} = diag\left(\left[e^{j\varphi_1}, e^{j\varphi_2}, \dots, e^{j\varphi_{N_{IRS}}}\right]^T\right),$$
$$\mathbf{W}^{opt} = \mathbf{U}_1. \quad (33)$$

A. EXHAUSTIVE SEARCH SCHEME

In terms of the optimization problem (33), the number of alternative matrix $\mathbf{\Phi}$ is finite since the phase shift of each reflecting element is discrete and is selected from the predefined phase shift set $\mathcal{F}$, where $\varphi_n \in \mathcal{F}, \forall n = 1, \cdots, N_{IRS}$. Under such a condition, the non-convex optimization problem (33) is able to be converted to a phase shift search problem whose optimal solution can be obtained by an exhaustive search method. Specifically, the exhaustive search method demands to successively generate the phase shift value $\{\varphi_n\}_{n=1}^{N}$ from $\mathcal{F}$, and then the phase shift matrix $\mathbf{\Phi}$ is able to be determined. Then, with a given $\mathbf{\Phi}$, the cascaded channel $\bar{\mathbf{H}}_e$ is definite and the corresponding combining matrix $\mathbf{W}$ can be obtained by utilizing the SVD operation of the cascaded channel $\bar{\mathbf{H}}_e$. Once the phase shift and the combining matrix is determined, the optimal data rate for the IRS-aided THz communication system can be got by computing all the possible candidates of $\mathbf{\Phi}$ and $\mathbf{W}$. After traversing

all the alternative matrices $\boldsymbol{\Phi}$ and the corresponding matrices $\boldsymbol{W}$, we are able to get the optimal communication rate and the expected phase shift matrix, denoted as $\boldsymbol{\Phi}^{opt}$. During this procedure, there are total $|\mathcal{F}|^{N_{IRS}}$ alternative phase shift matrices to participate in calculations. Therefore, the exhaustive search method suffers from severe computational complexity that involves a huge search operations for possible matrices $\boldsymbol{\Phi}$. To decrease the computational complexity, it is pressing to put forward the low complexity methods in practical application scenarios.

B. LOCAL SEARCH SCHEME

To begin with, the main complexity of the exhaustive search method comes from the exponential growth of the possible phase shift combinations when the number of the IRS elements increases. Attractively, different from the exhaustive search method, the proposed local search method can greatly reduce the complexity by just selecting the partially better combinations of discrete phase shifts. Similar to the exhaustive search method, the local search method transforms the optimization problem (33) into the phase shift search problem. The local search method just achieves the suboptimal data rate performance while it can greatly decrease the computational complexity. According to [59], the elements of the diagonal matrix $\{\boldsymbol{\Lambda}\,(i,i)\}_{i=1}^{N_s}$ are regarded as the virtual path gains. and the phase shift matrix $\boldsymbol{\Phi}$ is influenced by the virtual path gains. Instead, the combiner $\boldsymbol{W}$ is related with the unitary matrix $\boldsymbol{U}_1$ rather than the virtual path gains. Subsequently, we turn to maximize the sum of these virtual path gains rather than directly solving the non-convex optimization problem (33). Therefore, the problem (33) can be further transformed as

$$\max_{\boldsymbol{\Phi}} \log_2 \sum_{i=1}^{N_s} |\boldsymbol{\Lambda}\,(i,i)|^2 \overset{(a)}{=} \max_{\boldsymbol{\Phi}} \left[1 + Tr\left(\bar{\boldsymbol{H}}_e \bar{\boldsymbol{H}}_e^H \right) \right], \quad (34)$$

where the definition (a) comes from the linear algebra that for any matrix $\boldsymbol{A} \in \mathcal{C}^{N_s \times N_s}$ and $Tr\,(\boldsymbol{A}) \in \sum_{i=1}^{N_s} \lambda_i(\boldsymbol{A})$, we can get that $\{|\boldsymbol{\Lambda}(i,i)|^2\}_{i=1}^{N_s}$ is the eigenvalue of the term $\bar{\boldsymbol{H}}_e \bar{\boldsymbol{H}}_e^H$. To simplify the expression in (34), a more explicit formulation is employed to represent the problem as

$$\max_{\boldsymbol{\Phi}} Tr\left(\bar{\boldsymbol{H}}_e \bar{\boldsymbol{H}}_e^H \right)$$
$$s.t.\ \varphi_n \in \mathcal{F}, \quad \forall n = 1, \ldots, N_{IRS},$$
$$\boldsymbol{\Phi} = diag\left(\left[e^{j\varphi_1}, e^{j\varphi_2}, \ldots, e^{j\varphi_{N_{IRS}}} \right]^T \right). \quad (35)$$

From the problem (35) we note that this optimization problem is a phase shift search problem, in which just contains one matrix variable $\boldsymbol{\Phi}$. Compared with exhaustive search method, the proposed local search method owns some obvious advantages. On the one hand, by analyzing the structure features of the matrix $\bar{\boldsymbol{H}}_e$, the matrix variable $\boldsymbol{W}$ can be eliminated and the obtained problem (35) is not involved with SVD operations. On the other hand, the phase shift search strategy is optimized in the local search method. Instead of

calculating all the phase shift combinations, the proposed local search method only takes advantage of the relatively better combinations of the discrete phase shifts. Thus, based on these superiorities, the proposed local search method is capable of lessening the computational complexity tremendously by avoiding the exhaustive search of the phase shift combinations.

Algorithm 2 Local Search Algorithm

Require: BS-IRS channel: $\bar{\boldsymbol{H}}_1$, IRS-MS channel: $\bar{\boldsymbol{H}}_2$, the bit quantization number: b, the controllable phase shift range of each reflecting element: $\Delta\varphi$.
1: Initialize: Phase shift matrix: $\boldsymbol{\Phi} = \boldsymbol{0}_{N_{IRS} \times N_{IRS}}$, the phase shift set: $\mathcal{F} = \{0, \Delta\varphi/2^b, \cdots, (2^b - 1)\Delta\varphi/2^b\}$, $\gamma = Tr\left(\bar{\boldsymbol{H}}_e \bar{\boldsymbol{H}}_e^H \right)$, the maximum value of γ: $\gamma^{\max} = 0$.
   Iterative Process:
2: **for** $i = 1 : N_{IRS}$ **do**
3:    Randomly generate $\varphi_k \in \mathcal{F}, \forall k \in [i+1, \cdots, N_{IRS}]$,
4:    **for all** $\varphi_i \in \mathcal{F}$ **do**
5:      Construct the phase shift matrix of IRS:
      $\boldsymbol{\Phi} = diag\left[e^{j\varphi_1}, \cdots, e^{j\varphi_{N_{IRS}}} \right]$,
6:      Calculate the term γ with a given $\boldsymbol{\Phi}$:
      $\gamma = Tr\left(\bar{\boldsymbol{H}}_e \bar{\boldsymbol{H}}_e^H \right) = Tr\left(\left(\bar{\boldsymbol{H}}_2 \boldsymbol{\Phi} \bar{\boldsymbol{H}}_1 \right) \left(\bar{\boldsymbol{H}}_2 \boldsymbol{\Phi} \bar{\boldsymbol{H}}_1 \right)^H \right)$,
7:      **if** $\gamma > \gamma^{\max}$ **do**
8:        $\gamma^{\max} = \gamma$, $\boldsymbol{\Phi}^{opt} = \boldsymbol{\Phi}$,
9:      **end if**
10:    **end for**
11: **end for**
12: Calculate the optimal cascaded channel:
    $\bar{\boldsymbol{H}}_e^{opt} = \bar{\boldsymbol{H}}_2 \boldsymbol{\Phi}^{opt} \bar{\boldsymbol{H}}_1$,
13: Derive the optimal combiner $\boldsymbol{W}^{opt}$:
    $\bar{\boldsymbol{H}}_e = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^H, \boldsymbol{U} = [\boldsymbol{U}_1, \boldsymbol{U}_2], \boldsymbol{W}^{opt} = \boldsymbol{U}_1$,
14: Compute the maximum data rate $R^{\max}$ based on (31),
Ensure: $\boldsymbol{\Phi}^{opt}, \boldsymbol{W}^{opt}, R^{\max}$.

The detailed procedure of the proposed local search algorithm for the data-rate maximization problem is illustrated in Algorithm 2. Here, the important steps of the proposed algorithm are introduced in the following. Firstly, we initialize some key parameters that are utilized in the iterative process, such as the phase shift matrix, the phase shift set for each IRS element and so on. Next, the iterative process is to be described in detail. During the ith iteration process, we fix the phase shifts of reflecting elements from $i+1$ to N_{IRS}, and select the optimal phase shift from $\mathcal{F}$ for the ith reflecting element by maximizing the term $Tr\left(\bar{\boldsymbol{H}}_e \bar{\boldsymbol{H}}_e^H \right)$. Continue such iteration process for N_{IRS} times, we are able to find the optimal phase shift matrix $\boldsymbol{\Phi}^{opt}$, the optimal combiner $\boldsymbol{W}^{opt}$ and the corresponding data-rate $R^{\max}$ in the end.

C. DEEP NEURAL NETWORK BASED SCHEME

Through the analysis of optimization problem as well as the hardware constraints, intriguingly, we note that the phase shift matrix is related to inherent channel features, such as path

gain, AoAs and AoDs. In other words, there is a nonlinear functional relationship between the phase shifts of IRS elements and the estimated CSI. However, such an implicit relationship is hard to be described mathematically by the IRS mapping function. Thus, the DNN based method, which maps from the estimated CSI to the phase shift matrix, is an appropriate choice to fit the desired mapping function [60]. The local search method discussed above not only endures obvious performance loss, but also suffers from high computational complexity due to the enormous SVD operations and the matrix multiplication. Therefore, another advantage of our developed method is that the DNN is able to further mitigate the computational burden of the phase shift search problem by fitting the IRS mapping function instead of searching the phase shift combinations of reflecting elements. To this end, the DNN based method is developed in this subsection.
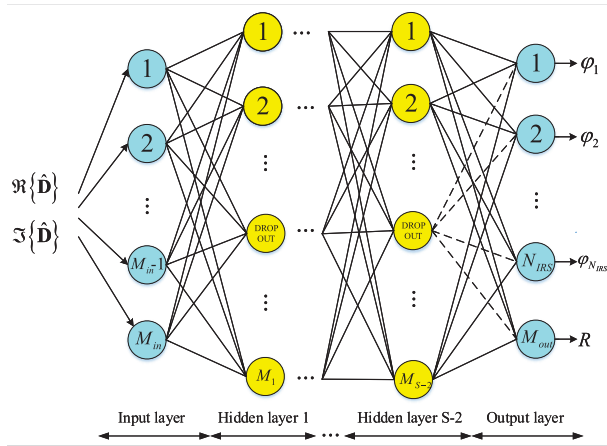

FIGURE 5. The proposed neural network architecture with *S* layers for the IRS-assisted THz communication system.

From [61] we know that DNN based method well fits the functional relationship when the training samples of the data sets is sufficient enough. Since the estimated CSI in previous section can provide enough data sets, the proposed DNN based method is able to tackle the optimization problem (33) with the low complexity and slight performance penalty compared with the exhaustive search method. Once the DNN based search model is established, the optimal phase shift matrix Φ as well as the corresponding communication rate R can be obtained quickly for the new CSI input. In the following, the neural network architecture of the proposed DNN base method is introduced, as shown in Fig. 5. The feedforward fully connected network is designed here, which contains an input layer, $S - 2$ hidden layers and an output layer. Also, any two neural units in different layers are interconnected with each other.

In the following, the proposed DNN based phase shift search algorithm is presented in detail. Generally, the proposed DNN based method consists of two distinct phases, namely (I) the offline (training) phase and (II) the online phase.

1) OFFLINE PHASE
In the training phase, the database can be generated by the channel estimation method mentioned above and the exhaustive search method, where the exhaustive search procedure operated online involves tremendous latency overhead. The estimated CSI provides the channel features while the exhaustive search method can find the mapping relationship between the optimal phase shift matrix (optimal data rate) and the channel features. Thus, the input vector $\bar{D}$ contains the path gain, AoAs and AoDs. Correspondingly, the output vector $\bar{O} = (\varphi_1, \cdots, \varphi_{N_{IRS}}, R)$ is the optimal phase shifts of IRS elements and the desired data rate. By aggregating the input data and the output data together, a complete database is created to train the proposed DNN model. Once the DNN model is trained well, the optimal phase shift matrix Φ can be gained with much lower computational complexity compared with the local search method and the exhaustive search method. It is worth noting that the collected channel samples are generated dynamically during the offline training process. Although the channel varies in practice, the trained DNN model still possesses the extremely high adaptability.

2) ONLINE PHASE
In the online training phase, we suppose that both the database and DNN model are placed on the IRS side. Before the DNN model comes into use, we firstly estimate a new channel sample. With the given CSI, the IRS feeds the preprocessing channel information to the trained DNN model. Finally, the optimal $R^{\max}$ and Φ^{opt} can be gained through the online processing of the proposed DNN model. We note that the online prediction only involves simple calculation operations in the neural network, such as the multiplication operation, and thus the proposed DNN based method has the better ability to reduced the computational complexity significantly.

As discussed above, the DNN, as one of the branches of deep learning [60]–[62], has an advantage in fitting the mapping functions between the phase shift matrix and the CSI. With this motivation, the S-layer DNN model is designed to seek the appropriate mapping functions. The critical parts of the DNN model are introduced concretely in the following.

a: FEATURES SELECTION
Through the analysis of optimization objective and the training samples, the array response vectors and the path gains of channel information are finally selected as the input features in the adopted DNN model. In the IRS-assisted THz MIMO communication model, we assume that the BS-IRS channel and IRS-MS channel contain L_1 and L_2 propagation paths, respectively. Let $\bar{a}_{BS}(\phi_{l_1}), \bar{a}_{IRS}(\gamma_{l_1}^{AOA}), \bar{a}_{IRS}(\gamma_{l_2}^{AOD}), \bar{a}_{MS}(\theta_{l_2})$ represent the estimated array response vectors. Meanwhile, the $\bar{\alpha}_{l_1}$ and $\bar{\beta}_{l_2}$ are the estimated path gains, where $l_1 \in (1, 2, \cdots, L_1)$ and $l_2 \in (1, 2, \cdots, L_2)$.

b: DATA PREPROCESSING
Before feeding the channel samples to the proposed DNN model, the input vector $\bar{D}$ needs to be normalized to conclude

the uniform statistical distribution of the training samples. In practice, since the THz wave endures serious path loss and molecular absorption loss, the propagation path gains of THz signals changes widely. Thus, the normalization process is indispensable to narrow the gap among the different values of the data samples and accelerate the convergence of the DNN model. After the data preprocessing of the original data sets, the processed data samples can be regarded as the effective input of the DNN model. Next, the formulation of the processed input vector $\bar{D}$ can be represented as

$$\hat{D} = \frac{1}{\bar{D}_{\max} - \bar{D}_{\min}} \bar{D} \tag{36}$$

where $D_{\max}$ and $D_{\min}$ are the maximum and minimum elements in $\left| \bar{D} \right|$.

c: INPUT

For a specific communication scenario, once the number of input features is determined, the dimension of the input layer is also definite. Although the input features of DNN have been selected according to the database, another challenge is that the input vector $\hat{D}$ of the proposed DNN has to be the real numbers. To face this difficulty, the complex input vector is essential to be divided into the real part and the imaginary part. Then, the real part and the imaginary part are integrated together to form the final input vector. For the sake of simplifying this explanation, we select the array response $\bar{a}_{BS}(\phi_{l_1})$ as an example. Specifically, the vector $\bar{a}_{BS}(\phi_{l_1})$ of path l_1 is split into the real parts $\Re(\bar{a}_{BS}(\phi_{l_1}))$ and the imaginary parts $\Im(\bar{a}_{BS}(\phi_{l_1}))$. Similarly, all the feature vectors can be split into real and imaginary parts. Therefor, the dimension of the expected input vector is $(N_{BS}L_1 + N_{IRS}L_1 + N_{IRS}L_2 + N_{MS}L_2 + L_1 + L_2)$ in the proposed DNN model, and the input vector can be further expressed as

$$\Re\left\{ \hat{D} \right\} = \Re \left\{ \begin{array}{c} \bar{a}_{BS}(\phi_{l_1}), \cdots, \bar{a}_{BS}(\phi_{L_1}) \\ \bar{a}_{IRS}(\gamma_{l_1}^{AOA}), \cdots, \bar{a}_{IRS}(\gamma_{L_1}^{AOA}) \\ \bar{a}_{IRS}(\gamma_{l_2}^{AOD}), \cdots, \bar{a}_{IRS}(\gamma_{L_2}^{AOD}) \\ \bar{a}_{MS}(\theta_{l_2}), \cdots, \bar{a}_{MS}(\theta_{L_2}) \\ \bar{\alpha}_1, \bar{\alpha}_2, \cdots, \bar{\alpha}_{L_1} \\ \bar{\beta}_1, \bar{\beta}_2, \cdots, \bar{\beta}_{L_2} \end{array} \right\} \tag{37}$$

$$\Im\left\{ \hat{D} \right\} = \Im \left\{ \begin{array}{c} \bar{a}_{BS}(\phi_{l_1}), \cdots, \bar{a}_{BS}(\phi_{L_1}) \\ \bar{a}_{IRS}(\gamma_{l_1}^{AOA}), \cdots, \bar{a}_{IRS}(\gamma_{L_1}^{AOA}) \\ \bar{a}_{IRS}(\gamma_{l_2}^{AOD}), \cdots, \bar{a}_{IRS}(\gamma_{L_2}^{AOD}) \\ \bar{a}_{MS}(\theta_{l_2}), \cdots, \bar{a}_{MS}(\theta_{L_2}) \\ \bar{\alpha}_1, \bar{\alpha}_2, \cdots, \bar{\alpha}_{L_1} \\ \bar{\beta}_1, \bar{\beta}_2, \cdots, \bar{\beta}_{L_2} \end{array} \right\} \tag{38}$$

d: OUTPUT

It is worth mentioning that the data-rate (R) is correlated with the optimal phase shift matrix Φ, so the proposed DNN model can output the R and the diagonal elements $\varphi = (\varphi_1, \cdots, \varphi_{N_{IRS}})$ of the matrix Φ simultaneously. Thus, the final output vector is represented as $\bar{O} = (\varphi_1, \cdots, \varphi_{N_{IRS}}, R)$.

TABLE 1. Complexity comparisons of considered schemes.

Algorithm	Computational Complexity
Exhaustive search	$\mathcal{O}\left(N_{MS}^3 N_{BS}^2 N_s N_{IRS}^{2^b} \right)$
Local search	$\mathcal{O}\left(N_{MS}^2 N_{BS} N_{IRS} 2^b \right)$
DNN	$\mathcal{O}\left(\begin{array}{c} (N_{BS} + 2N_{IRS} + N_{MS} + 2L)Ln_2 \\ + \sum_{i=2}^{S-2} n_i n_{i+1} + n_{S-1}(N_{IRS} + 1) \end{array} \right)$

e: TRAINING PROCESS OF DNN

On the basis of the working principle of the neural network, each current neural unit sums all the input values which are related to the neural units from the previous layer and the corresponding weights. More specifically, the input value for each current neural unit sums the product of the output values from the previous layer and the corresponding weights. In the hidden layers, the summation values need to be processed by rectified linear units (ReLUs) to improve nonlinear relations and alleviate gradient dissipation problems. Then, the output values of current neural units can be further utilized for the next layer. To avoid the overfitting issue, the dropout units are also designed in the proposed neural network architecture, where the dropout units indicate that a random neural unit is selected as an inactive unit during the learning process.

In order to obtain the optimal Φ and R with small batches of training samples as well as the low computational complexity, the stochastic gradient descent (SGD) algorithm is employed in the proposed neural network architecture. According to the SGD algorithm, the parameters of the DNN model at kth layer can be expressed as

$$\Psi^k = \Psi^{k-1} - \eta \nabla \mathcal{L}\left(\Psi^{k-1} \right) \tag{39}$$

where η is the learning rate of the algorithm, Ψ is the set of all the neural network parameters, and $\nabla \mathcal{L}(\cdot)$ is the gradient of the batch's loss function. After that all the parameters are determined, the DNN model is established accordingly.

According to the trained DNN model, when a new data sample is input, the predicted output $\hat{O}$ can be obtained swiftly. To demonstrate the effectiveness of the proposed DNN based method, the minimum mean square error (MMSE) criterion is chosen to minimize the loss function $L(\Psi)$. Also, the MMSE can evaluate the prediction accuracy of the proposed DNN model by calculating the difference between the predicted output $\hat{O}$ and the sampled output $\bar{O}$. During the training process of the DNN model, we need to minimize the regression loss function until the proposed DNN converges. So, the loss function $L(\Psi)$ is expressed as

$$L(\Psi) = MMSE(\hat{O} - \bar{O}) \tag{40}$$

D. COMPLEXITY ANALYSIS OF DIVERSE SCHEMES

In this subsection, the detailed complexity of different phase shift search schemes are analyzed, including the exhaustive search scheme, local search scheme, and DNN based scheme for the IRS aided THz communication scenarios.

To begin with, the computational complexity of the exhaustive search method mainly caused by three different aspects. Firstly, calculating the multiplication of two matrix

$\bar{\boldsymbol{H}}_e \in \mathcal{C}^{N_{MS} \times N_{BS}}$ and $\bar{\boldsymbol{H}}_e^H \in \mathcal{C}^{N_{BS} \times N_{MS}}$ is $\mathcal{O}(N_{MS}N_{BS}N_{MS})$. Then, the complexity of searching all the possible phase shift matrices is $\mathcal{O}(N_{IRS}^{2^b})$. In addition, the complexity of the SVD operation about the matrix $\bar{\boldsymbol{H}}_e$ is $\mathcal{O}(N_{MS}N_{BS}N_s)$. Therefore, the total complexity of the exhaustive search scheme is

$$\mathcal{O}(N_{MS}N_{BS}N_s)\,\mathcal{O}\left(N_{IRS}^{2^b}\right)\mathcal{O}(N_{MS}N_{BS}N_{MS}) \\ = \mathcal{O}\left(N_{MS}^3 N_{BS}^2 N_s N_{IRS}^{2^b}\right). \quad (41)$$

Next, since the local search scheme only needs to search $N_{IRS}2^b$ possible phase shift combinations, the complexity of the local search scheme can be written as

$$\mathcal{O}(N_{MS}N_{BS}N_{MS})\,\mathcal{O}\left(N_{IRS}2^b\right) \\ = \mathcal{O}\left(N_{MS}^2 N_{BS}N_{IRS}2^b\right). \quad (42)$$

Without loss of generality, the complexity of the offline training can be ignored, the major computational complexity comes from the online phase. Based on the algorithm procedure of the DNN based model, the complexity of online predictions mainly involves the operations of multiplying the output value of each unit with the corresponding weight. Hence, the complexity between two layers is the product of the input vector and the weight matrix. In terms of the S-lay DNN model, the total complexity of DNN based scheme can be determined as

$$\mathcal{O}\left(\begin{array}{c} (N_{BS} + 2N_{IRS} + N_{MS} + 2L)Ln_2 \\ + \sum_{i=2}^{S-2} n_i n_{i+1} + n_{S-1}(N_{IRS} + 1) \end{array}\right) \quad (43)$$

where $n_i \in (n_1, n_2 \ldots n_S)$ is the number of neural units for the ith layer. Especially for the $S = 4$, the complexity of such a DNN model can be expressed as

$$\mathcal{O}\left(\begin{array}{c} (N_{BS} + 2N_{IRS} + N_{MS} + 2L)\,Ln_2 \\ + n_2 n_3 + n_3\,(N_{IRS} + 1) \end{array}\right) \quad (44)$$

V. NUMERICAL RESULTS
In this section, numerical results are presented to demonstrate the performance of the IRS-aided THz MIMO system and the conventional THz communication system without IRS in terms of the data rate performance and the computational complexity. Here, the conventional scheme without IRS means that all the system settings are the same as the proposed IRS assisted system except for the existence of IRS, and thus we can see the benefits brought by the IRS. For a given downlink communication scenario, the BS employs $N_{BS} = 32$ antennas and the MS uses $N_{MS} = 32$ antennas. As for the IRS, we adopt $N_{IRS} = 16$ reflecting elements with 2 bits quantization precision and the working frequency of the IRS elements is at 0.22 THz. Particularly, the distance between adjacent antenna elements or the neighbouring reflecting elements is set as the half the wavelength of the THz wave. Due to the extremely sparse nature of the THz channel $\bar{\boldsymbol{H}}_1$ and $\bar{\boldsymbol{H}}_2$, we denote that the number of the propagation paths is $L_1 = L_3 = 3$. Additionally, the complex gains are selected from the circularly-symmetric Gaussian distribution and satisfies $\bar{\alpha}_{l_1}, \bar{\beta}_{l_2} \in \mathcal{CN}(0, 1)$, where

$l_1, l_2 \in [1, 2, 3]$. On the basis of the geometric channel model, the physical AoAs and AoDs are randomly generated from $[0, 2\pi]$ with an equal probability. Ultimately, it worth noting that the channel parameters are estimated in previous section and thus are known here.
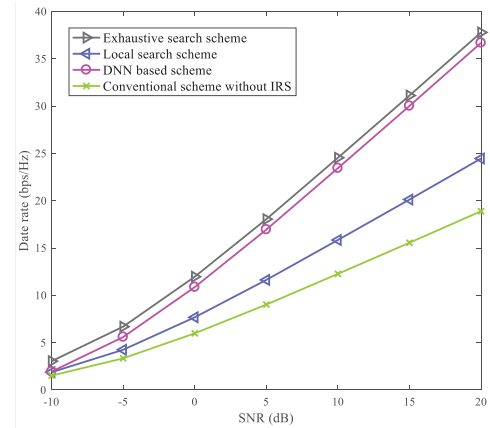

FIGURE 6. Data rate performance comparisons of considered schemes versus different values of SNR.

A. DATA RATE PERFORMANCE COMPARISONS
Fig. 6 depicts the data-rate performance comparisons of diverse phase shift search methods along with the increase of the SNR. In the practical THz communication scenarios, it is obvious that all the proposed phase shift search schemes with IRS achieves higher data rate performance than the conventional scheme without IRS. Among these proposed phase shift search schemes, the exhaustive search method performs the optimal performance while the proposed local search method suffers from the apparent performance penalty. Specifically, when the SNR is 10 dB, the exhaustive search method is about 7.59 bps/Hz higher than the local search method. Interestingly, the simulation curve of the proposed DNN based method can be basically consistent with the curve of the optimal method (around 1.14 bps/Hz performance gap at SNR = 10 dB). One may also note that the DNN based method is able to realize the better communication performance compared with the proposed local search method in indoor THz communication systems.

Fig. 7 illustrates the data rate performance versus the number of reflecting elements N_{IRS} among different schemes under the condition that SNR = 10 dB. From Fig. 7 we can see that, the data rate performance of the conventional scheme without IRS remains unchanged when the number of the reflecting elements increases. Fortunately, the number of IRS elements has a great impact on the communication performance of our proposed schemes, including the exhaustive search scheme, the local search scheme and the DNN based scheme. In other words, by increasing the number of the IRS elements, all the proposed schemes is able to enhance the data rate performance significantly. Specifically, in the case that the number of IRS elements is configured as 60, the data rate of the proposed DNN based scheme is 38.36 bps/Hz, which is greatly better than that of local search
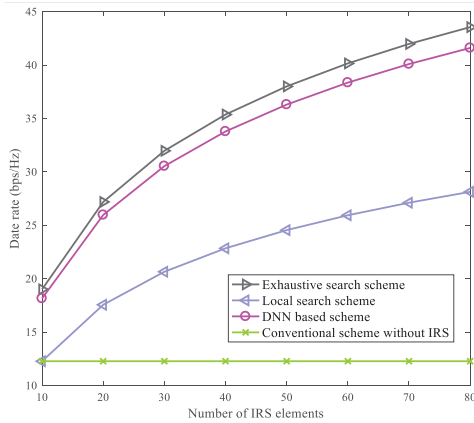
FIGURE 7. Data rate performance comparisons of considered schemes versus the number of reflecting elements with SNR = 10 dB.
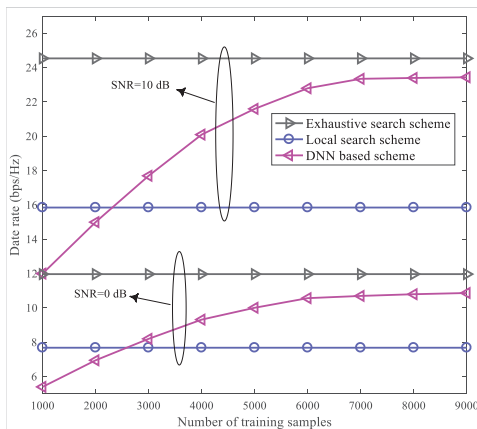

FIGURE 8. Data rate performance comparisons of considered schemes along with the number of training samples.

method with 12.24 bps/Hz. Moreover, it is worth mentioning that the data rate performance of the proposed DNN based scheme approaches the simulation curve of the exhaustive search scheme.

Fig. 8 presents the data rate performance of different scheme versus the number of training samples. It is obvious that the proposed local search method and the exhaustive search method are not influenced by the number of training samples. Nevertheless, the data rate of the proposed DNN based method becomes continuously better along with the increasing number of the training samples, and has a tendency to approach the the exhaustive search method. Notably, when the number of the training samples is 8000, the DNN based method has almost converged in terms of SNR = 5 dB and SNR = 10 dB. Under the case that the training samples are sufficient at SRN = 5 dB, the DNN based method inherits about 1.17 bps/Hz performance loss compared with the exhaustive search method, and possesses about 3.12 bps/Hz performance enhancement compared with the local search method. In short, as long as the database is large enough, the DNN based method can attain the near optimal data rate performance for the IRS aided THz communication systems.
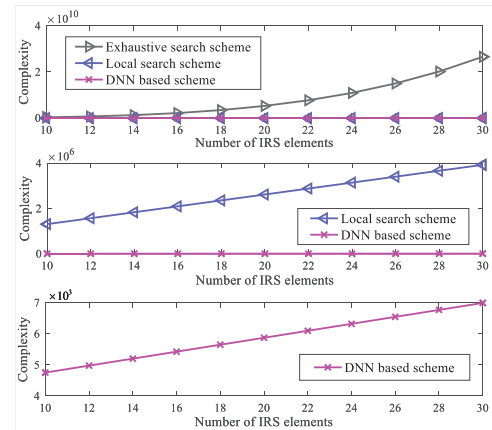

FIGURE 9. Complexity comparisons of diverse schemes versus the number of reflecting elements N_{IRS}.

B. COMPLEXITY COMPARISONS

Fig. 9 investigates the computational complexity comparisons of the proposed three different schemes with the increasing of the IRS elements. For the sake of a fair comparison condition, the proposed diverse phase shift search methods are based on the same parameter settings just as mentioned before. Especially, the proposed DNN model contains four lays where $n_2 = 16$ and $n_3 = 16$. Firstly, from Fig. 9 we find that the proposed local search method can greatly reduce the computational overhead in comparison with the exhaustive search method. Moreover, the further complexity reduction is able to be achieved dramatically by the proposed DNN based method compared with the local search method. In more concrete terms, when the number of the IRS elements is 16, the complexity of the exhaustive search, the local search method and the DNN based method is 2.15×10^9, 2.10×10^6 and 5.42×10^3, respectively. Generally, the complexity of offline phase is not taken into account here. Additionally, one may point out that, as the number of the reflecting elements increases, the computational complexity of both the local search method and the exhaustive search method grows to a great extent, while the proposed DNN based method varies slowly. Thus, in other words, the advantage of the proposed DNN based method method in terms of the complexity tends to be more distinct. Ultimately, combining the Fig. 6 and Fig. 9, the proposed DNN based scheme is capable of making a better trade-off between the data rate performance and the computational complexity, which will be widely deployed in future IRS assisted THz communication scenarios.

VI. CONCLUSION

THz frequency band is regarded as a promising alternative to provide large bandwidth and high data rates for 6G wireless communication systems. To settle the poor coverage capability and high path loss of THz waves existing in indoor THz communication systems, the IRS has been proposed recently to enhance the communication rates of the terminal equipment as well as the coverage capacity. By utilizing the CS technique, we put forward a low complexity

IAP-SP scheme to solve the channel estimation problem that involves the double grid quantization issue. Fortunately, the proposed IAP-SP algorithm achieves the basically same recovery performance compared with classical SP algorithm. With the estimated CSI, we then propose three diverse data rate maximization methods by searching the phase shifts of IRS elements. Remarkably, the proposed DNN based scheme makes a better tradeoff between computational complexity and data rate performance. Specifically, as long as the training samples are large enough, the proposed DNN-based scheme is able to achieve the near-optimal performance and debases the computational burden sufficiently. It is also worth noting that these proposed frameworks in this paper are able to be extended to multiuser MIMO communication scenarios. Ultimately, the IRS-assisted THz communication systems will act an important role in indoor 6G application scenarios.

APPENDIX
A. PROOF OF (23)
According to (22), the vectorized channel is written as

$$
\begin{aligned}
&vec\left(\boldsymbol{H}\right) \\
&= vec\left(\bar{\boldsymbol{A}}_{MS}\boldsymbol{G}_2\bar{\boldsymbol{A}}_{IRS,2}^H\boldsymbol{\Phi}\bar{\boldsymbol{A}}_{IRS,1}\boldsymbol{G}_1\bar{\boldsymbol{A}}_{BS}^H\right) \\
&= \left(\bar{\boldsymbol{A}}_{BS}^*\otimes\bar{\boldsymbol{A}}_{MS}\right)vec\left(\boldsymbol{G}_2\bar{\boldsymbol{A}}_{IRS,2}^H\boldsymbol{\Phi}\bar{\boldsymbol{A}}_{IRS,1}\boldsymbol{G}_1\right) \\
&= \underbrace{\left(\bar{\boldsymbol{A}}_{BS}^*\otimes\bar{\boldsymbol{A}}_{MS}\right)}_{Part1}\underbrace{\left(\boldsymbol{G}_1^T\otimes\boldsymbol{G}_2\right)}_{Part2}\underbrace{vec\left(\bar{\boldsymbol{A}}_{IRS,2}^H\boldsymbol{\Phi}\bar{\boldsymbol{A}}_{IRS,1}\right)}_{part3}. \quad (45)
\end{aligned}
$$

Then, we continue to vectorize (45) and can obtain the following expression as

$$
\begin{aligned}
vec\left(\boldsymbol{H}\right) &= vec\left(vec\left(\boldsymbol{H}\right)\right) \\
&= \left[vec\left(\bar{\boldsymbol{A}}_{IRS,2}^H\boldsymbol{\Phi}\bar{\boldsymbol{A}}_{IRS,1}\right)\right]^T \\
&\quad \otimes\left(\bar{\boldsymbol{A}}_{BS}^*\otimes\bar{\boldsymbol{A}}_{MS}\right)vec\left(\boldsymbol{G}_1^T\otimes\boldsymbol{G}_2\right) \\
&= \left[\bar{\boldsymbol{A}}_{IRS,1}^T\otimes\bar{\boldsymbol{A}}_{IRS,2}^Hvec\left(\boldsymbol{\Phi}\right)\right]^T \\
&\quad \otimes\left(\bar{\boldsymbol{A}}_{BS}^*\otimes\bar{\boldsymbol{A}}_{MS}\right)vec\left(\boldsymbol{G}_1^T\otimes\boldsymbol{G}_2\right) \\
&= \left[\left(vec\left(\boldsymbol{\Phi}\right)\right)^T\left(\bar{\boldsymbol{A}}_{IRS,1}\otimes\bar{\boldsymbol{A}}_{IRS,2}^*\right)\right] \\
&\quad \otimes\left(\bar{\boldsymbol{A}}_{BS}^*\otimes\bar{\boldsymbol{A}}_{MS}\right)vec\left(\boldsymbol{G}_1^T\otimes\boldsymbol{G}_2\right). \quad (46)
\end{aligned}
$$

Thus, by substituting (46)) into (20), we accomplish the proof for (23).

B. PROOF OF (28)
Considering the (27), we note that

$$
\begin{aligned}
\bar{\boldsymbol{G}}_v &= vec^{-1}\left(\bar{\boldsymbol{g}}_v\right) = \bar{\boldsymbol{G}}_1^T\otimes\bar{\boldsymbol{G}}_2 \\
&= \begin{bmatrix} \bar{g}_{1,1}^1\otimes\bar{\boldsymbol{G}}_2 & \cdots & \bar{g}_{1,N_1}^1\otimes\bar{\boldsymbol{G}}_2 \\ \vdots & \ddots & \vdots \\ \bar{g}_{N_1,1}^1\otimes\bar{\boldsymbol{G}}_2 & \cdots & \bar{g}_{N_1,N_1}^1\otimes\bar{\boldsymbol{G}}_2 \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{A}_{1,1} & \cdots & \boldsymbol{A}_{1,N_1} \\ \vdots & \ddots & \vdots \\ \boldsymbol{A}_{N_1,1} & \cdots & \boldsymbol{A}_{N_1,N_1} \end{bmatrix}, \quad (47)
\end{aligned}
$$

where $\boldsymbol{A}_{i_1,j_1} = \bar{g}_{i_1,j_1}^1\otimes\bar{\boldsymbol{G}}_2, \forall i_1, j_1 \in [1, N_1]$.

Based on the fact that $\bar{g}_{i_1,j_1}^1$ is a complex number, so we can get

$$
\begin{aligned}
\sum_{i_1=1}^{N_1}\sum_{j_1=1}^{N_1}\boldsymbol{A}_{i_1,j_1} &= \sum_{i_1=1}^{N_1}\sum_{j_1=1}^{N_1}\bar{g}_{i_1,j_1}^1\otimes\bar{\boldsymbol{G}}_2 \\
&= \bar{\boldsymbol{G}}_2\sum_{i_1=1}^{N_1}\sum_{j_1=1}^{N_1}\bar{g}_{i_1,j_1}^1 = C_0\bar{\boldsymbol{G}}_2. \quad (48)
\end{aligned}
$$

Thus, the proof for (28) is completed.

C. PROOF OF (29)
On the basic of (47), the entry of $\bar{\boldsymbol{G}}_1$ can be formulated as

$$
\bar{g}_{i_1,j_1}^1 = \frac{\sum_{i_2=1}^{N_2}\sum_{j_2=1}^{N_2}a_{i_1,j_1}\left(i_2,j_2\right)}{\sum_{i_2=1}^{N_2}\sum_{j_2=1}^{N_2}\bar{g}_{i_2,j_2}^2} \quad (49)
$$

where $\bar{g}_{i_2,j_2}^2 \in \bar{\boldsymbol{G}}_2$, $a_{i_1,j_1}\left(i_2,j_2\right) \in \boldsymbol{A}_{i_1,j_1}, \forall i_1, j_1 \in [1, N_1]$, $\forall i_2, j_2 \in [1, N_2]$. By resorting to (49), the proof for (29) can be completed directly.

REFERENCES
[1] Z. Chen, X. Ma, B. Zhang, Y. X. Zhang, Z. Niu, N. Kuang, W. Chen, L. Li, and S. Li, "A survey on terahertz communications," *China Commun.*, vol. 16, no. 2, pp. 1–35, Feb. 2019.
[2] H.-J. Song and T. Nagatsuma, "Present and future of terahertz communications," *IEEE Trans. Terahertz Sci. Technol.*, vol. 1, no. 1, pp. 256–263, Sep. 2011.
[3] I. F. Akyildiz, J. M. Jornet, and C. Han, "Terahertz band: Next frontier for wireless communications," *Phys. Commun.*, vol. 12, pp. 16–32, Sep. 2014.
[4] J. Federici and L. Moeller, "Review of terahertz and subterahertz wireless communications," *J. Appl. Phys.*, vol. 107, no. 11, Jun. 2010, Art. no. 111101.
[5] S. Mollahasani and E. Onur, "Evaluation of terahertz channel in data centers," in *Proc. IEEE/IFIP Netw. Oper. Manage. Symp. (NOMS)*, Apr. 2016, pp. 727–730.
[6] S. U. Hwu, K. B. de Silva, and C. T. Jih, "Terahertz (THz) wireless systems for space applications," in *Proc. IEEE Sensors Appl. Symp.*, Feb. 2013, pp. 171–175.
[7] I. F. Akyildiz, C. Han, and S. Nie, "Combating the distance problem in the millimeter wave and terahertz frequency bands," *IEEE Commun. Mag.*, vol. 56, no. 6, pp. 102–108, Jun. 2018.
[8] X. Ma, Z. Chen, Z. Li, W. Chen, and K. Liu, "Low complexity beam selection scheme for terahertz systems: A machine learning approach," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2019, pp. 1–6.
[9] T. J. Cui, M. Q. Qi, X. Wan, J. Zhao, and Q. Cheng, "Coding metamaterials, digital metamaterials and programmable metamaterials," *Light: Sci. Appl.*, vol. 3, no. 10, p. e218, Oct. 2014.
[10] C. Liaskos, S. Nie, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. Akyildiz, "A new wireless communication paradigm through software-controlled metasurfaces," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 162–169, Sep. 2018.
[11] W. Chen, X. Ma, Z. Li, and N. Kuang, "Sum-rate maximization for intelligent reflecting surface based terahertz communication systems," in *Proc. IEEE/CIC Int. Conf. Commun. Workshops China (ICCC Workshops)*, Aug. 2019, pp. 153–157.
[12] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, Aug. 2019.
[13] C. Huang, G. C. Alexandropoulos, A. Zappone, M. Debbah, and C. Yuen, "Energy efficient multi-user MISO communication using low resolution large intelligent surfaces," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2018, pp. 1–6.

[14] 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.

[15] J. Zhang, E. Björnson, M. Matthaiou, D. W. K. Ng, H. Yang, and D. J. Love, "Prospective multiple antenna technologies for beyond 5G," 2019, *arXiv:1910.00092*. [Online]. Available: http://arxiv.org/abs/1910.00092

[16] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," 2019, *arXiv:1905.00152*. [Online]. Available: http://arxiv.org/abs/1905.00152

[17] C. Tienda, M. Arrebola, J. A. Encinar, and G. Toso, "Analysis of a dual-reflect array antenna," *IET Microw., Antennas Propag.*, vol. 5, no. 13, pp. 1636–1645, Oct. 2011.

[18] H. Giddens, L. Yang, J. Tian, and Y. Hao, "Mid-infrared reflect-array antenna with beam switching enabled by continuous graphene layer," *IEEE Photon. Technol. Lett.*, vol. 30, no. 8, pp. 748–751, Apr. 15, 2018.

[19] F. Gao, T. Cui, and A. Nallanathan, "On channel estimation and optimal training design for amplify and forward relay networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1907–1916, May 2008.

[20] R. Zhang, Y.-C. Liang, C. C. Chai, and S. Cui, "Optimal beamforming for two-way multi-antenna relay channel with analogue network coding," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 5, pp. 699–712, Jun. 2009.

[21] H. Bolcskei, R. U. Nabar, O. Oyman, and A. J. Paulraj, "Capacity scaling laws in MIMO relay networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1433–1444, Jun. 2006.

[22] K. Han and K. Huang, "Wirelessly powered backscatter communication networks: Modeling, coverage, and capacity," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2548–2561, Apr. 2017.

[23] G. Wang, F. Gao, R. Fan, and C. Tellambura, "Ambient backscatter communication systems: Detection and performance analysis," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4836–4846, Nov. 2016.

[24] C. Boyer and S. Roy, "Backscatter communication and RFID: Coding, energy, and MIMO analysis," *IEEE Trans. Commun.*, vol. 62, no. 3, pp. 770–785, Mar. 2014.

[25] L. Subrt and P. Pechac, "Intelligent walls as autonomous parts of smart indoor environments," *IET Commun.*, vol. 6, no. 8, pp. 1004–1010, May 2012.

[26] L. Subrt and P. Pechac, "Controlling propagation environments using intelligent walls," in *Proc. 6th Eur. Conf. Antennas Propag. (EUCAP)*, Mar. 2012, pp. 1–5.

[27] L. Subrt, P. Pechac, A. Vallecchi, L. Ford, R. Langley, and J. Rigelsford, "Advanced modelling of intelligent walls for indoor smart environments," in *Proc. 8th Eur. Conf. Antennas Propag. (EuCAP)*, Apr. 2014, pp. 620–622.

[28] X. Tan, Z. Sun, J. M. Jornet, and D. Pados, "Increasing indoor spectrum sharing capacity using smart reflect-array," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–6.

[29] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, Nov. 2019.

[30] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network: Joint active and passive beamforming design," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–6.

[31] X. Ma, Z. Chen, W. Chen, Y. Chi, Z. Li, C. Han, and Q. Wen, "Intelligent reflecting surface enhanced indoor terahertz communication systems," *Nano Commun. Netw.*, vol. 24, May 2020, Art. no. 100284.

[32] Q. Wu and R. Zhang, "Beamforming optimization for intelligent reflecting surface with discrete phase shifts," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 7830–7833.

[33] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling large intelligent surfaces with compressive sensing and deep learning," 2019, *arXiv:1904.10136*. [Online]. Available: http://arxiv.org/abs/1904.10136

[34] D. Mishra and H. Johansson, "Channel estimation and low-complexity beamforming design for passive intelligent surface assisted MISO wireless energy transfer," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 4659–4663.

[35] A. A. M. Saleh and R. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE J. Sel. Areas Commun.*, vol. 5, no. 2, pp. 128–137, Feb. 1987.

[36] S. Priebe, C. Jastrow, M. Jacob, T. Kleine-Ostmann, T. Schrader, and T. Kurner, "Channel and propagation measurements at 300 GHz," *IEEE Trans. Antennas Propag.*, vol. 59, no. 5, pp. 1688–1698, May 2011.

[37] R. Piesiewicz, C. Jansen, D. Mittleman, T. Kleine-Ostmann, M. Koch, and T. Kurner, "Scattering analysis for the modeling of THz communication systems," *IEEE Trans. Antennas Propag.*, vol. 55, no. 11, pp. 3002–3009, Nov. 2007.

[38] Z. Gao, L. Dai, S. Han, C.-L. I, Z. Wang, and L. Hanzo, "Compressive sensing techniques for next-generation wireless communications," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 144–153, Jun. 2018.

[39] 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.

[40] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.

[41] C. L. Holloway, E. F. Kuester, J. A. Gordon, J. O'Hara, J. Booth, and D. R. Smith, "An overview of the theory and applications of metasurfaces: The two-dimensional equivalents of metamaterials," *IEEE Antennas Propag. Mag.*, vol. 54, no. 2, pp. 10–35, Apr. 2012.

[42] B. O. Zhu, J. Zhao, and Y. Feng, "Active impedance metasurface with full 360° reflection phase tuning," *Sci. Rep.*, vol. 3, p. 3059, Oct. 2013.

[43] W.-Z. Xu, F.-F. Ren, J. Ye, H. Lu, L. Liang, X. Huang, M. Liu, I. V. Shadrivov, D. A. Powell, G. Yu, B. Jin, R. Zhang, Y. Zheng, H. H. Tan, and C. Jagadish, "Electrically tunable terahertz metamaterials with embedded large-area transparent thin-film transistor arrays," *Sci. Rep.*, vol. 6, no. 1, p. 23486, Sep. 2016.

[44] A. H. C. Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and A. K. Geim, "The electronic properties of graphene," *Phys. Status Solidi*, vol. 244, no. 11, pp. 4106–4111, 2007.

[45] A. Andryieuski and A. V. Lavrinenko, "Graphene metamaterials based tunable terahertz absorber: Effective surface conductivity approach," *Opt. Express*, vol. 21, no. 7, pp. 9144–9155, Apr. 2013.

[46] Y. Zhang, Y. Feng, B. Zhu, J. Zhao, and T. Jiang, "Graphene based tunable metamaterial absorber and polarization modulation in terahertz frequency," *Opt. Express*, vol. 22, no. 19, pp. 22743–22752, Sep. 2014.

[47] S. H. Lee, M. Choi, T.-T. Kim, S. Lee, M. Liu, X. Yin, H. K. Choi, S. S. Lee, C.-G. Choi, S.-Y. Choi, X. Zhang, and B. Min, "Switching terahertz waves with gate-controlled active graphene metamaterials," *Nature Mater.*, vol. 11, no. 11, pp. 936–941, Nov. 2012.

[48] E. Carrasco, M. Tamagnone, and J. Perruisseau-Carrier, "Tunable graphene reflective cells for THz reflectarrays and generalized law of reflection," *Appl. Phys. Lett.*, vol. 102, no. 10, Mar. 2013, Art. no. 104103.

[49] 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.

[50] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Hybrid precoding for millimeter wave cellular systems with partial channel knowledge," in *Proc. Inf. Theory Appl. Workshop (ITA)*, Feb. 2013, pp. 1–5.

[51] C. Han, A. O. Bicen, and I. F. Akyildiz, "Multi-ray channel modeling and wideband characterization for wireless communications in the terahertz band," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2402–2412, May 2015.

[52] J. Lee, G.-T. Gil, and Y. H. Lee, "Exploiting spatial sparsity for estimating channels of hybrid MIMO systems in millimeter wave communications," in *Proc. IEEE Global Commun. Conf.*, Dec. 2014, pp. 3326–3331.

[53] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[54] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4680–4688, Jul. 2011.

[55] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[56] R. E. Carrillo and K. E. Barner, "Lorentzian iterative hard thresholding: Robust compressed sensing with prior information," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4822–4833, Oct. 2013.

[57] Z. Wan, Z. Gao, B. Shim, K. Yang, G. Mao, and M.-S. Alouini, "Compressive sensing based channel estimation for millimeter-wave full-dimensional MIMO with lens-array," *IEEE Trans. Veh. Technol.*, vol. 69, no. 2, pp. 2337–2342, Feb. 2020.

[58] K. Ying, Z. Gao, S. Lyu, Y. Wu, H. Wang, and M.-S. Alouini, "GMD-based hybrid beamforming for large reconfigurable intelligent surface assisted millimeter-wave massive MIMO," *IEEE Access*, vol. 8, pp. 19530–19539, 2020.

[59] B. Ning, Z. Chen, W. Chen, and J. Fang, "Beamforming optimization for intelligent reflecting surface assisted MIMO: A Sum-Path-Gain maximization approach," 2019, *arXiv:1909.07282*. [Online]. Available: http://arxiv.org/abs/1909.07282

[60] K. Gurney, *An Introduction to Neural Networks*. Boca Raton, FL, USA: CRC Press, 2014.

[61] L. Deng and D. Yu, "Deep learning: Methods and applications," *Found. Trends signal Process.*, vol. 7, nos. 3–4, pp. 197–387, 2013.

[62] V. Sze, Y.-H. Chen, T.-J. Yang, and J. S. Emer, "Efficient processing of deep neural networks: A tutorial and survey," *Proc. IEEE*, vol. 105, no. 12, pp. 2295–2329, Dec. 2017.

YAOJIA CHI received the B.S. degree in electronic science and technology from Lanzhou Jiaotong University, Lanzhou, China, in 2018. She is currently pursuing the M.S. degree with the School of Electronic Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China. Her current research interests include terahertz wave control and intelligent reflecting surface hardware design and modulation.

XINYING MA received the B.S. degree from Southwest Jiaotong University, Chengdu, China, in 2015, and the M.S. degree from the University of Electronic Science and Technology of China (UESTC), Chengdu, in 2018, where he is currently pursuing the Ph.D. degree with the National Key Laboratory of Science and Technology on Communications. His current research interests include terahertz (THz) communications, millimetre-wave communications, intelligent reflecting surface (IRS), compressed sensing (CS), and machine learning.

ZHI CHEN (Senior Member, IEEE) received the B.E., M.E., and Ph.D. degrees in electrical engineering from the University of Electronic Science and Technology of China (UESTC), in 1997, 2000, and 2006, respectively. In April 2006, he joined the National Key Laboratory of Science and Technology on Communications, UESTC. He was a Visiting Scholar with the University of California, Riverside, from 2010 to 2011. He has been a Professor with UESTC, since 2013. His current research interests include 5G mobile communications, the tactile Internet, and terahertz communication. He has served as a Reviewer for various international journals and conferences, including the IEEE Transactions on Vehicular Technology and the IEEE Transactions on Signal Processing.

CHONG HAN (Member, IEEE) received the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, USA, in 2016. He is currently a tenure-track Associate Professor with the University of Michigan-Shanghai Jiao Tong University (UM-SJTU) Joint Institute, Shanghai Jiao Tong University, China. His research interests include terahertz band, millimeter-wave communication networks, and electromagnetic nanonetworks. He is a member of ACM. He was a recipient of the 2017 Shanghai Sailing Program, the 2018 Elsevier Nano Communication Network Journal (NanoComNet) Young Investigator Award, and the 2018 Shanghai Chenguang Program. He is a TPC Chair to organize the multiple IEEE and ACM conferences and workshops. He has been an Editor with *Nano Communication Network Journal* (Elsevier), since 2016. He has been an Associate Editor with IEEE Access, since 2017.

WENJIE CHEN received the B.E. degree in communication engineering from Southwest Jiaotong University (SWJTU), Chengdu, China, in 2018. He is currently pursuing the M.E. degree in communication with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu. His research interests include massive multiple-input multiple-output and intelligent reflecting surface (IRS).

SHAOQIAN LI (Fellow, IEEE) received the B.E. degree in communication technology from the Northwest Institute of Telecommunication Engineering (Xidian University), Xi'an, China, in 1981, and the M.E. degree in information and communication systems from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 1984.

In 1984, he joined UESTC as an Academic Member, where he has been a Professor of information and communication systems, since 1997, and also been a Ph.D. Supervisor, since 2000. He is currently the Director of the National Key Laboratory of Science and Technology on Communications, UESTC. He holds over 60 granted and filed patents. His general interests include wireless and mobile communications, anti-jamming technologies, and signal processing for communications subjects, on which he has authored or coauthored over 100 journal articles, 100 conference papers, and two edited books. His current research interests include multiple-antenna signal processing technologies for mobile communications, cognitive radios, and coding and modulation for next-generation mobile broadband communications systems. He has been a member of the Communication Expert Group of the National 863 Plan, since 1998, and has been the Future Project, since 2005. He has served the various IEEE conferences as the Technical Program Committee (TPC) Member. He serves as a member for the Board of Communications and Information Systems, the Academic Degrees Committee, the State Council of China, and the Expert Group of Key Special-Project on Next-Generation Broadband Wireless Mobile Communications of China (approved by the State Council, since 2007). He was a TPC Co-Chair of the 2005, 2006, and the 2008 IEEE International Conference on Communications, Circuits, and Systems. He is also a member of the Editorial Board of *Chinese Science Bulletin* and *Chinese Journal of Radio Science*.

ZHUOXUN LI received the B.E. degree in communication engineering from the University of Electronic Science and Technology of China (UESTC), in 2018, where he is currently pursuing the M.E. degree with the National Key Laboratory of Science and Technology on Communications. His research and study interests include terahertz communication and intelligent reflecting surface (IRS).