Beamforming Technologies for Ultra-Massive MIMO in Terahertz Communications

Boyu Ning, Member, IEEE, Zhongbao Tian, Student Member, IEEE,
Weidong Mei, Member, IEEE, Zhi Chen, Senior Member, IEEE, Chong Han, Member, IEEE,
Shaoqian Li, Fellow, IEEE, Jinhong Yuan, Fellow, IEEE and Rui Zhang, Fellow, IEEE

Abstract—Terahertz (THz) communications with a frequency band of 0.1–10 THz are envisioned as a promising solution to future high-speed wireless communication. Although with tens of gigahertz available bandwidth, THz signals suffer from severe free-spreading loss and molecular-absorption loss, which limit the wireless transmission distance. To compensate for the propagation loss, the ultra-massive multiple-input-multiple-output (UM-MIMO) can be applied to generate a high-gain directional beam by beamforming technologies. In this paper, a review of beamforming technologies for THz UM-MIMO systems is provided. Specifically, we first present the system model of THz UM-MIMO and identify its channel parameters and architecture types. Then, we illustrate the basic principles of beamforming via UM-MIMO and discuss the far-field and near-field assumptions in THz UM-MIMO. Moreover, an important beamforming strategy in THz band, i.e., beam training, is introduced wherein the beam training protocol and codebook design approaches are summarized. The intelligent-reflecting-surface (IRS)-assisted joint beamforming and multi-user beamforming in THz UM-MIMO systems are studied, respectively. The spatial-wideband effect and frequency-wideband effect in the THz beamforming are analyzed and the corresponding solutions are provided. Further, we present the corresponding fabrication techniques and illuminate the emerging applications benefiting from THz beamforming. Open challenges and future research directions on THz UM-MIMO systems are finally highlighted.

Index Terms—Terahertz communications, ultra-massive MIMO, terahertz channel model, wideband beamforming, multi-user MIMO, intelligent reflecting surface, terahertz antenna array.

I. INTRODUCTION

Since the beginning of the 21st century, the evolution of online applications on social networks has led to an unprecedented growing number of wireless subscribers, who require real-time connectivity and tremendous data consumption [1], [2]. In this backdrop, various organizations and institutions have published their standards to support the wireless traffic explosion, such as the 3rd generation partnership project (3GPP) long term evolution (LTE) [3], the wireless personal area network (WPAN) [4], the wireless high definition (WiHD) [5], IEEE 802.15 [6], and IEEE 802.11 wireless local area network (WLAN) [7]. During the evolution of these standards, one common feature is that the consumed frequency spectrum increased steadily to satisfy the dramatic demands for instantaneous information. According to Edholm’s Law of Bandwidth [8], the telecommunications data rate doubles every 18 months. The wireless data rate requirement is expected to meet 100 Gbps to fulfill different growing service requirements before 2030.

The emerging millimeter-wave (mmWave)-band communications in the fifth-generation (5G) standard are able to achieve considerable improvements in the network capacity, with the vision of meeting demands beyond the capacity of previous-generation systems. Although some advanced means like the multiple-antenna technologies [9], the coordinated multi-point (CoMP) technologies [10], and the carrier aggregation (CA) technologies [11] may boost the data rate reaching several Gbps in mmWave communications, the achievable data rate on each link will be significantly suppressed when the number of connected users and/or devices is growing. To further raise the throughput, enhance the spectral efficiencies, and increase the connections, academic and industrial research proposed to deploy a large number of low-power small cells giving rise to the heterogeneous networks (HetNets) [12], [13]. Meanwhile, novel multiple access schemes, such as orthogonal time-frequency space (OTFS) [14], [15], rate-splitting [16], [17], and non-orthogonal multiple access (NOMA) [18], [19] technologies have been investigated for meeting the above requirements by sharing the resource block, e.g., a time slot, a frequency channel, a spreading code, or a spatial beam. However, due to the frequency regulations, current mmWave communications have limited available bandwidth and at present, there is even no frequency block wider than 10 GHz left unoccupied in the bands below 100 GHz.

To alleviate the spectrum bottleneck and realize at least 100 Gbps communications in the future, new spectral bands should be explored to support the data-hungry applications, e.g., virtual reality (VR) and augmented reality (AR), which require microsecond latency and ultra-fast download. To this end, the terahertz (THz) band (0.1–10 THz) has received noticeable attention in the research community as a promising candidate...
for various scenarios with high-speed transmission. As shown in Fig. 1, the bands below and above THz band have been extensively explored, including radio waves, microwave/mmWave, and free-space optical (FSO). In terms of signal generation, the THz band is exactly between the frequency regions generated by oscillator-based electronic and emitter-based photonic approaches, which incurs difficulty of electromagnetic generation, known as “the last piece of radio frequency (RF) spectrum puzzle for communication systems” [20]. In terms of wireless transmission, the most urgent challenges lie in the physical-layer hindrance, i.e., high spreading loss and severe molecular absorption loss in THz electromagnetic propagation.

Multiple-input-multiple-output (MIMO) systems, which can generate high-gain directional beams via beamforming technologies, are considered a pragmatic solution to compensate for the propagation loss. For metallic-antenna arrays, the element spacing is commonly half wavelength. With the increase of frequency adopted in wireless systems, more elements can be packed in the same size of the antenna array. Thus, compared to the MIMO in microwave systems, the concept of massive MIMO with tens to hundreds of elements in mmWave systems has been raised in [21]. In 2016, Ian F. Akyildiz and Josep Miquel Jornet introduced the concept of ultra-massive MIMO (UM-MIMO) in THz systems [22], which showed that the number of elements of the metamaterial-based or graphene-based nano-antenna THz array can be much larger than that of the mmWave massive MIMO. This is because the antenna spacing of nano-antenna arrays can be much smaller than half wavelength, meanwhile, the wavelength of the THz band is smaller than its mmWave counterpart.

In this paper, we reuse the term “UM-MIMO” to describe the THz antenna array with hundreds to thousands of elements, which is not specific to nano-antenna array (as in [22]) but compatible with conventional half-wavelength metallic-antenna array. We start with the fundamental concepts related to the THz UM-MIMO and precisely illuminate the principle of beamforming via affluent graphs, which are significant and instructive for communication researchers or engineers. Moreover, this paper covers some on-trend research points and classifies the existing THz MIMO arrays, which sheds light on the current research status and the progress of THz UM-MIMO. Finally, the emerging applications and open challenges are elaborated. This paper serves as a review to provide useful technical guidance and inspiration for future research and implementation of beamforming in THz UM-MIMO systems.

A. Related Tutorials, Magazines, and Surveys

Despite that the THz technology is not as mature as microwave/mmWave and FSO, the gap is progressively being closed thanks to its rapid development in recent years. Table I summarizes the representative works in this field including tutorials, magazines, and surveys published in the last two decades.

The first survey was conducted by P. H. Siegel, in which existing THz applications, sensors, and sources are presented [23]. In 2004, M. J. Fitch and R. Osiander further discussed the sources, detectors, and modulators for practical THz systems [24]. In 2007, several magazines came up with reports of THz technology progress status and applications of THz systems [25–27]. In 2010, J. Federici and L. Moeller provided the first overview focusing on the THz communications including channel coding, generation methods, detection, antennas, and link measurements [28]. After that, K. C. Huang and Z. Wang provided a tutorial on constructing robust, low-cost THz wireless systems, in 2011 [29]. In the same year, T. Nagatsuma et al. discussed the current progress of THz communications applications and highlighted some issues that need to be considered for the future of THz systems [30–32]. In 2012, K. Wu et al. provided a tutorial on THz antenna technologies, and T. Kurner et al. reported the standardization at IEEE 802.15 IG THz [33, 34]. In 2014, T. Kurner and S. Priebe reported the current research projects, spectrum regulations, and ongoing standardization activities in THz communication systems [35]. I. F. Akyildiz et al. reported the state-of-the-art THz technologies and highlighted the challenges from the communication and networking perspective as well as in terms of experimental testbeds [39]. In 2015, A. Hirata and M. Yaita gave a brief overview of the THz technologies and standardization of wireless communications [40]. In 2016, C. Lin and G. Y. Li reported an array-of-subarrays structure for THz wireless systems and discussed the benefits in terms of circuit and communication [41]. M. Hasan et al. provided an overview of the progress on graphene-based devices and Nagatsuma, T. et al. gave a tutorial on the photonics technologies in THz communications [42, 43]. J. F. Federici et al. gave a survey to illustrate the impact of weather on THz wireless links [44]. In 2017, S. Muntaz et al. discussed the opportunities and challenges in THz communications for vehicular networks [45]. In 2018, V. Petrov et al. provided a tutorial on propagation modeling, antenna, and testbed designs [46]. A. Boulogeorgos et al. reported the basic system architecture for THz wireless links with bandwidths into optical networks [47]. C. Han and Y. Chen reported three methods, e.g., deterministic, statistical, and hybrid methods, to model THz propagation channels [48]. I. F. Akyildiz et al. focused on the solution to the THz distance limitation [49]. N. Khalid et al. provided a tutorial on performing THz modulation schemes [50]. D. Headland [51] provided a tutorial on the basic principles of beam control in the THz band.

Since 2019, the number of papers on THz communications has increased notably. There were two magazines and five surveys that contain the development progress, unresolved problems, latest solutions, standardization works, and opportunities...
### TABLE I: Tutorials, magazines, and surveys on THz communications

| Authors                          | Year | Title                                                      | Type   | Brief Description                                                                                                                                 |
|----------------------------------|------|------------------------------------------------------------|--------|---------------------------------------------------------------------------------------------------------------------------------------------------|
| P. H. Siegel                     | 2002 | Terahertz technology                                       | Survey | This paper gives an overview of THz technology applications, sensors, and sources, with some discussion on science drivers, historical background, and future trends. |
| M. J. Fitch and R. Osiander      | 2004 | Terahertz waves for communications and sensing            | Magazine| This article reports the THz technology for communications and sensing applications. Sources, detectors, and modulators are also discussed for practical systems. |
| R. Pieszewicz. et al.            | 2007 | Short-range ultra-broadband terahertz communications: concepts and perspectives | Magazine| This article reports the concept of ultra-broadband THz communication and gives the potential applications of such a system supporting multi-gigabit data rates. |
| I. Hosako et al.                 | 2007 | At the dawn of a new era in terahertz technology          | Magazine| This article reports the developments in these fields such as THz quantum cascade lasers, THz quantum well photodetectors, an ultra-wideband THz time domain spectroscopy system, an example of a database for materials of fine art, and results from measuring atmospheric propagation. |
| M. Tonouchi                      | 2007 | Cutting-edge terahertz technology                         | Magazine| This article reports the THz technology progress status and expected usages in wireless communication, agriculture, and medical applications. |
| J. Federici and L. Moeller       | 2010 | Review of terahertz and subterahertz wireless communications | Survey | This paper gives an overview of THz communication systems, which demonstrate basic channel coding, generation methods, detection, antennas, and link measurements. |
| K. c. Huang and Z. Wang          | 2011 | Terahertz terabit wireless communication                  | Tutorial| This paper provides a tutorial to construct robust, low-cost wireless systems for THz terabit communications. |
| T. Kleine-Ostmann and T. Nagatsuma | 2011 | A review on terahertz communications research            | Magazine| This article reports the emerging technologies and system researches that might lead to ubiquitous THz communication systems in the future. |
| T. Nagatsuma                     | 2011 | Terahertz technologies: present and future                | Magazine| This article reports the latest progress in THz technologies in sources, detectors, and system applications. Future challenges toward market development are also discussed. |
| H. J. Song and T. Nagatsuma      | 2011 | Present and future of terahertz communications            | Survey | This paper gives an overview of the current progress of THz communications applications and discusses some issues that need to be considered for the future of THz systems. |
| K. Wu et al.                     | 2012 | Substrate-integrated millimeter-wave and terahertz antenna technology | Tutorial| This paper provides a tutorial on mmWave and THz antenna technologies including the planar/nonplanar antenna structures and provides a promising technological platform for mmWave and THz wireless systems. |
| T. Kurner et al.                 | 2012 | Towards future terahertz communications systems          | Magazine| This article reports the technology development, demonstrations of data transmission, ongoing activities in standardization at IEEE 802.15 IG THz, and the regulation of the spectrum beyond 300 GHz. |
| T. Nagatsuma et al.              | 2013 | Terahertz wireless communications based on photonics technologies | Tutorial| This paper provides a tutorial on recent works on THz wireless communications systems based on photonic signal generation at carrier frequencies of over 100 GHz. |
| K. Jha, G. Singh                 | 2013 | Terahertz planar antennas for future wireless communication: A technical review | Magazine| This paper gives an overview of high directivity antennas, high-power sources, and efficient detectors for compact THz communication systems, with a special focus on improving planar antenna gain by reducing conductor and substrate losses. |
| I. F. Akyildiz et al.            | 2014 | Terahertz band: Next frontier for wireless communications  | Magazine| This article reports the THz applications and challenges in the generation, channel modeling, and communication systems, along with a brief discussion on experimental and simulation testbeds. |
| T. Kurner and S. Priebe          | 2014 | Towards THz communications - Status in Research, standardization and regulation | Magazine| This article reports the current research projects, spectrum regulations, and ongoing standardization activities in THz communication systems. |
| I. F. Akyildiz et al.            | 2014 | TeraNets: ultra-broadband communication networks in the terahertz band | Magazine| This article reports the state of the art in THz Band device technologies and highlights the challenges and potential solutions from the communication and networking perspective as well as in terms of experimental testbeds. |
| A. Hirata and M. Yaita           | 2015 | Ultrafast terahertz wireless communications technologies | Survey | This paper gives an overview of the development of the THz technologies and standardization of wireless communications. |
| Authors                  | Year | Title                                                                                      | Type   | Brief Description                                                                                                                                                                                                 |
|-------------------------|------|--------------------------------------------------------------------------------------------|--------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| C. Lin and G. Y. Li     | 2016 | Terahertz communications: An array-of-subarrays solution                                   | Magazine | This article reports the indoor multi-user THz communication systems with antenna arrays and discusses how the array-of-subarrays structure benefits THz communications from both the circuit and communication perspectives. |
| M. Hasan et al.         | 2016 | Graphene terahertz devices for communications applications                                  | Survey  | This paper gives an overview of recent progress on graphene-based devices for modulation, detection, and generation of THz waves, which are among the key components for future THz band communications systems. |
| T. Nagatsuma et al.     | 2016 | Advances in terahertz communications accelerated by photonics                               | Tutorial| This paper provides a tutorial on the latest trends in THz communications research, focusing on how photonics technologies have played a key role in the development of first-age THz communication systems. |
| J. F. Federici et al.   | 2016 | Review of weather impact on outdoor terahertz wireless communication links                  | Survey  | This paper gives an overview of the impact of weather on THz wireless links and emphasizes THz attenuation and channel impairments caused by atmospheric gases, airborne particulates, refractive index inhomogeneities, and their associated scintillations. |
| S. Muntaz et al.        | 2017 | Terahertz communication for vehicular networks                                            | Survey  | This paper gives an overview of the opportunities and challenges in THz communications for vehicular networks.                                                                                                    |
| V. Petrov et al.        | 2018 | Last meter indoor terahertz wireless access: Performance insights and implementation roadmap | Magazine | This article reports the propagation modeling, antenna, and testbed designs, along with a step-by-step roadmap for THz Ethernet extension for indoor environments.                                                                 |
| A. A. A. Boulougeorgos et al. | 2018 | Terahertz technologies to deliver optical network quality of experience in wireless systems beyond 5G | Magazine | This article reports the basic system architecture for THz wireless links with bandwidths of more than 50 GHz into optical networks.                                                                                     |
| C. Han and Y. Chen      | 2018 | Propagation modeling for wireless communications in the terahertz band                     | Magazine | This article reports the channel modeling in the THz band, based on the deterministic, statistical, and hybrid methods. The state-of-the-art THz channel models in single-antenna and UM-MIMO systems are extensively reviewed, respectively. |
| I. F. Akyildiz et al.   | 2018 | Combating the distance problem in the millimeter wave and terahertz frequency bands        | Magazine | This article reports the research advances on physical layer distance adaptive design, UM-MIMO, reflectarrays, and hypersurfaces to show the direction to solve the problem of THz limited transmission distance. |
| N. Khalid et al.        | 2018 | Energy-efficient modulation and physical layer design for low terahertz band communication channel in 5G femtocell Internet of Things | Tutorial | This paper provides a tutorial on the modulation schemes, the hardware parameters, and the circuit blocks in the THz band which are suitable for mass market production. |
| D. Headland et al.      | 2018 | Tutorial: Terahertz beamforming, from concepts to realizations                             | Tutorial | This paper provides a tutorial on the basic principles of beam control in the THz range from the perspective of array antenna theory and diffraction optics and demonstrates significant demonstrations, including conventional optics, phased array antennas, leaky-wave antennas, and passive arrays. |
| Z. Chen et al.          | 2019 | A survey on terahertz communications                                                        | Survey  | This paper gives an overview of the development towards THz communications and presents some key technologies faced in THz wireless communication systems.                                                                 |
| K. Tekbıyık et al.      | 2019 | Terahertz band communication systems: Challenges, novelities and standardization efforts    | Survey  | This paper gives an overview of the unresolved problems, the latest solutions, and the standardization works in the THz communication systems.                                                                     |
| T. S. Rappaport et al.  | 2019 | Wireless communications and applications above 100 GHz: Opportunities and challenges for 6G and beyond | Survey  | This paper gives an overview of the technical challenges and opportunities for wireless communication and sensing applications above 100 GHz and presents several promising discoveries, novel approaches, and recent results. |
| K. K. O et al.          | 2019 | Opening terahertz for everyday applications                                                  | Magazine | This article reports the devices in CMOS, the challenges in implementing THz circuits, the performance of CMOS THz circuits, and their applications and expected advances.                                                          |
| K. M. S. Haq et al.     | 2019 | Terahertz-enabled wireless systems for beyond-5G ultra-fast networks: A brief survey        | Magazine | This article reports the applications utilizing THz bands and hints at future research directions in this rapidly developing area.                                                                                |
| X. Fu et al.            | 2020 | Terahertz beam steering technologies: From phased arrays to field-programmable metasurfaces | Survey  | This article gives a comprehensive summary of the principles and characteristics of THz beam steering technology from two aspects, the conventional technologies, and the reconfigurable metasurface-based technologies. |
| Authors                  | Year | Title                                                                                       | Type    | Brief Description                                                                                                                                                                                                                   |
|-------------------------|------|----------------------------------------------------------------------------------------------|---------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| H. Elavam et al. [70]   | 2020 | Terahertz Band: The last piece of RF spectrum puzzle for communication systems                 | Survey  | This paper gives an overview of the recent activities on the development, standardization, and applications in THz communications.                                                                                                      |
| H. Sariieddeen et al.   | 2020 | Next generation terahertz communications: A rendezvous of sensing, imaging, and localization   | Magazine| This article reports the THz technologies that bring significant advances to the areas of wireless communications, imaging, sensing, and localization.                                                                                |
| L. Zhang et al. [59]    | 2020 | Beyond 100 Gb/s optoelectronic terahertz communications: Key technologies and directions      | Magazine| This article reports the key technologies of optoelectronic THz communications in the physical layer, including approaches of broadband devices, baseband signal processing technologies, and the design of advanced transmission system architectures. |
| M. A. Jamshed et al.    | 2020 | Antenna selection and designing for THz applications: Suitability and performance evaluation: A survey | Survey  | This paper gives an overview of the characteristics of THz band, THz-enabled applications, materials of THz antenna, design parameters, and approaches to measure the performance of a THz-enabled antenna.                              |
| S. Ghafoor et al.       | 2020 | MAC Protocols for Terahertz Communication: A Comprehensive Survey                              | Survey  | This paper gives an overview of THz MAC protocols with classifications, band features, design issues, and future challenges.                                                                                                 |
| C. X. Wang et al.       | 2020 | 6G Wireless Channel Measurements and Models: Trends and Challenges                            | Magazine| This article reports the application scenarios, performance metrics, potential key technologies, and future research challenges of 6G wireless communication networks.                                                             |
| A. Faisal et al. [63]   | 2020 | Ultramassive MIMO systems at terahertz bands: Prospects and challenges                         | Magazine| This article reports recent advances in transceiver design and channel modeling and discusses the major challenges and shortcomings by deriving the relationships among communication range, array dimensions, and system performance.     |
| J. Tan et al. [64]      | 2020 | THz precoding for 6G: Applications, challenges, solutions, and opportunities                    | Magazine| This article reports three THz precoding architectures to cope with the challenges of severe path loss and limited coverage in THz communications.                                                                                       |
| F. Lemic et al. [65]    | 2021 | Survey on Terahertz Nanocommunication and Networking: A Top-Down Perspective                  | Survey  | This paper gives an overview of the current THz applications, different layers of the protocol stack, as well as the available channel models and experimentation tools.                                                                  |
| C. Han et al. [66]      | 2021 | Hybrid Beamforming for Terahertz Wireless Communications: Challenges, Architectures, and Open Problems | Magazine| This article reports the challenges and characteristics of THz hybrid beamforming design and compares different hybrid beamforming architectures for THz communications.                                                                |
| H. Sariieddeen et al.   | 2021 | An Overview of Signal Processing Techniques for Terahertz Communications                       | Survey  | This paper gives an overview of the recent developments in signal processing techniques for THz communications, focusing on waveform modulation, channel estimation, beamforming, and precoding in the THz UM-MIMO systems.               |
| H. J. Song [68]         | 2021 | Terahertz Wireless Communications: Recent Developments Including a Prototype System for Short-Range Data Downloading | Magazine| This article reports device technologies that can be employed for THz communication prototype systems, including photonic devices, Si-CMOS, SiGe HBTs, and III-V HEMTs/HBTs.                                                                 |
| H. Do et al. [69]       | 2021 | Terahertz line-of-sight MIMO communication: Theory and practical challenges                    | Magazine| This paper reports that MIMO spatial multiplexing in the THz band is feasible even under LoS conditions with reconfigurable array architectures, and provides insights from the information-theoretic perspective. |
| Z. Chen et al. [70]     | 2021 | Terahertz wireless communications for 2030 and beyond: A cutting-edge frontier                | Magazine| This paper reports four promising directions for future THz communications, including integrated sensing and communications, ultra-massive MIMO, and dynamic hybrid beamforming, intelligent surfaces.          |
| Z. Chen et al. [71]     | 2021 | Intelligent reflecting surface assisted terahertz communications toward 6G                    | Magazine| This paper reports the expectation and realization progress of incorporating the transformative features of the intelligent reflecting surface into THz wireless communications.                                                                      |
| C. Chaccour et al. [72] | 2022 | Seven Defining Features of Terahertz (THz) Wireless Systems: A Fellowship of Communication and Sensing | Survey  | This paper gives an overview of seven features of THz wireless systems, including THz band characteristics, tailored network architectures, synergy with lower frequencies, communication, and sensing integration, physical layer signal processing, multiple access, and network optimization. |
| C. Han et al. [73]      | 2022 | Terahertz Wireless Channels: A Holistic Survey on Measurement, Modeling, and Analysis        | Survey  | This paper provides a comprehensive overview of the study of THz wireless channels, and categorizes existing channel modeling methodologies into deterministic, stochastic, and hybrid approaches.                                                   |
for THz communications [20], [52]–[57]. In 2020, H. Sarieddeen et al. reported the current THz technologies in wireless communications, imaging, sensing, and localization [58]. L. Zhang et al. reported the key technologies of optoelectronic THz communications in the physical layer [59]. M. A. Jamshed et al. conducted a survey on the antenna selection design for THz applications [60]. S. Ghafoor et al. gave an overview of THz media access control (MAC) protocols with classifications, band features, design issues, and future challenges [61]. C. X. Wang et al. reported the channel measurements, characteristics, and models for THz frequency band [62]. A. Faisal et al. reported the advantages of UM-MIMO systems in THz communications and discussed the challenges and shortcomings [63]. In 2021, F. Lemic et al. conducted a survey on THz nano-communication and network from a top-down perspective [65]. C. Han et al. investigated the architectures and challenges of hybrid beamforming in THz communications [66]. H. Sarieddeen [67] gave an overview of signal processing techniques and propagation characteristics of THz wireless systems, respectively. H. J. Song et al. reported the device technologies employed for THz prototype systems [68]. H. Do et al. reported that spatial multiplexing is feasible even under line-of-sight (LoS) conditions with reconfigurable array architectures [69]. Z. Chen et al. reported promising directions for future THz communications [70], [71]. C. Chaccour et al. [72] gave an overview of seven features of THz wireless systems. C. Han et al. [73] provided a comprehensive overview of the study of THz wireless channels.

With the rapidly advancing of THz technologies in terms of new manufacturing materials, transceiver architectures, and antenna designs, THz UM-MIMO has been envisioned as a key paradigm for future wireless systems [74]. In the above-mentioned works, the authors in [63] provided a brief overview of THz UM-MIMO systems and highlighted a few challenges. The authors in [48] reviewed several channel models of THz UM-MIMO systems. However, to the best of our knowledge, there is still lacking a detailed tutorial overview tailored for THz UM-MIMO beamforming technologies, to cover their theoretical breakthroughs, novel technological developments, engineering fabricating issues, and practical deployment considerations. This tutorial thus aims to fill this gap by providing a holistic view of the THz UM-MIMO beamforming, including the basic principles, wideband effects, beamforming scenarios, existing MIMO arrays, emerging applications, and future challenges.

B. Contributions of this Paper

The main contributions of this paper are summarized below.

- We present a basic system model for THz UM-MIMO and propose a way to determine the channel parameters. The effects of the antenna geometry and transceiver architecture on the THz UM-MIMO systems are further discussed with precise definitions.
- We illustrate the basic principles of beamforming by visualizing the electromagnetic field distribution in the physical layer. We characterize the features of beam patterns generated by large-scale array and discuss the near-field and far-field assumptions in THz UM-MIMO systems.
- We introduce a beam training scheme that can realize the beamforming in THz UM-MIMO systems without using the channel state information (CSI). The basic concept of the beam training is elaborated and two essential components, i.e., training protocol and codebook design, are discussed. The lens antenna array is introduced to realize beam training as a special manner.
- We study an intelligent reflecting surface (IRS)-assisted system for THz communication, which helps enhance the signal coverage and improve spectral efficiency with very low energy consumption. The system model of the IRS-assisted THz UM-MIMO is presented and the joint active and passive beamforming strategies for THz communication are discussed.
- We introduce the multi-user UM-MIMO channels and provide an effective beamforming strategy for multi-user THz UM-MIMO systems. Based on this strategy, we revisit the popular linear beamforming algorithms that can eliminate

![Fig. 2. The structure of this paper.](image-url)
the inter-user interference and evaluate their performance in THz UM-MIMO systems.

- We present two main issues, i.e., spatial-wideband effect and frequency-wideband effect, which are notable and need to be considered in the wideband THz beamforming. We explain why these effects occur and provide some available solutions to addressing these issues.
- We provide an overview of the existing THz MIMO arrays based on different fabrication techniques, including electronic-based, photonics-based, and new materials-based. We pay special attention to those THz antenna arrays with dynamic beam steering capabilities.
- We identify the transformative functions of the THz beamforming technologies in emerging applications, including satellite communications, vehicle connectivity, indoor wireless networks, wireless data centers, secure transmission, and inter-chip communications. These applications demonstrate the great potential of THz UM-MIMO systems in shaping future wireless networks.
- We point out various open challenges faced by the THz UM-MIMO systems. In Section II, we introduce the THz UM-MIMO systems. Section III, we illuminate the basic beamforming principles. Sections IV and V present respectively the beam training and the IRS-assisted joint beamforming in THz UM-MIMO systems. Section VI and VII respectively discuss the multi-user beamforming schemes and wideband beamforming solutions in THz UM-MIMO systems. Section VIII provides an overview of the existing THz MIMO arrays. Section IX identifies the transformative functions of the THz beamforming technologies in emerging applications, including satellite communications, vehicle connectivity, indoor wireless networks, wireless data centers, secure transmission, and inter-chip communications. These applications demonstrate the great potential of THz UM-MIMO systems in shaping future wireless networks.
- We identify the open challenges and future research directions. As shown in Fig. 2, the rest of this paper is organized as follows. Section II introduces the THz UM-MIMO systems. In Section III, we illuminate the basic beamforming principles. Sections IV and V present respectively the beam training and the IRS-assisted joint beamforming in THz UM-MIMO systems. Section VI and VII respectively discuss the multi-user beamforming schemes and wideband beamforming solutions in THz UM-MIMO systems. Section VIII provides an overview of the existing THz MIMO arrays. Section IX identifies the transformative functions of the THz beamforming technologies in emerging applications, including satellite communications, vehicle connectivity, indoor wireless networks, wireless data centers, secure transmission, and inter-chip communications. These applications demonstrate the great potential of THz UM-MIMO systems in shaping future wireless networks. Finally, we conclude the paper in Section X. Finally, we conclude the paper in Section XI.

C. Organization and Notations

As shown in Fig. 2, the rest of this paper is organized as follows. Section II introduces the THz UM-MIMO systems. In Section III, we illuminate the basic beamforming principles. Sections IV and V present respectively the beam training and the IRS-assisted joint beamforming in THz UM-MIMO systems. Section VI and VII respectively discuss the multi-user beamforming schemes and wideband beamforming solutions in THz UM-MIMO systems. Section VIII provides an overview of the existing THz MIMO arrays. Section IX identifies the transformative functions of the THz beamforming technologies in emerging applications, including satellite communications, vehicle connectivity, indoor wireless networks, wireless data centers, secure transmission, and inter-chip communications. These applications demonstrate the great potential of THz UM-MIMO systems in shaping future wireless networks. Finally, we conclude the paper in Section X. Finally, we conclude the paper in Section XI.

Notation: We use small normal face for scalars, small bold face for vectors, and capital bold face for matrices. The superscript \{\cdot \}^T, \{\cdot \}^H, and \{\cdot \}^H denote the transpose, conjugate, and Hermitian transpose, respectively. \| \cdot \| and \| \cdot \|_F represent the modulus operator and Frobenius norm, respectively. \text{diag}(\cdot) denotes a diagonal matrix whose diagonal elements are given by its argument. \mathbb{R}, \mathbb{C} stand for the set of real numbers and the set of complex numbers, respectively.

II. THz UM-MIMO SYSTEMS

In this section, we start with describing the system model and specifying the parameters of THz UM-MIMO systems. Then, some important system characteristics are elaborated, i.e., array response vectors and transceiver architectures.

A. System Model

Consider a point-to-point THz UM-MIMO system with the quasi-static block fading channel. Let \(N_t\) and \(N_r\) denote the number of transmit and receive antenna elements, respectively. The received signal vector \(y \in \mathbb{C}^{N_r}\) can be expressed as

\[ y = Hx + n, \]

where \(x \in \mathbb{C}^{N_t}\) is the transmitted signal vector. \(H \in \mathbb{C}^{N_r \times N_t}\) is the channel matrix and \(n \in \mathbb{C}^{N_r}\) is the noise vector.

1) Channel Model: The channel matrix \(H\) can be modelled based on the channel measurements in THz band [75]–[79]. Important channel parameters include path loss exponent (PLE), shadow fading (SF), K factor (KF), delay spread (DS), azimuth spread of arrival (ASA), and elevation spread of arrival (ESA). These parameters in THz band are quite different from those in cmWave and mmWave scenarios, as shown in Table II [23]. Here, we present a popular statistical channel model, i.e., Saleh-Valenzuela (S-V) channel model [85], which is widely used in UM-MIMO systems. With the employment of ultra-massive antenna elements, the THz channel generally shows sparsity and strong directivity, which is composed of a strong LoS cluster and several weak NLoS clusters [86]–[88]. As such, the channel matrix in the time-domain can be written as

\[ H(t) = \sum_{i=1}^{N_{cl}} \sum_{j=1}^{N_{ray}} \delta(t - \tau_{ij}) \sqrt{\alpha_{ij}} G_{t,ij} G_{r,ij} a_{\tau_{ij}} a_{\tau_{ij}}^H, \]

where \(N_{cl}\) and \(N_{ray}\) are the number of clusters and the number of rays in the \(i\)-th cluster, respectively. \(\delta(\cdot)\) is the impulse response function. \(\alpha_{ij}\) denotes the path gain of the \(j\)-th ray in the \(i\)-th cluster. \(\tau_{ij} = T_i + T_{ij}\), in which \(T_i\) (with \(T_1 = 0\)) is

\[^{\dagger}\text{In particular, THz channel model can be classified as deterministic [80], [81], statistical [82], and hybrid model [83], [84]. Deterministic model requires detailed geometric knowledge of the propagation environment, which is site-specific and subject to high complexity. In comparison, statistical channel model allows fast channel construction based on key channel statistics.}\]
and $T_{ij}$ (with $T_{i1} = 0$) represent the arrival time of the $i$-th cluster and that of the $j$-th ray in it, respectively. $G_{t,ij}$ and $G_{r,ij}$ (resp. $a_{t,ij}$ and $a_{r,ij}$) represent the transmit and receive antenna gains (resp. array response vectors at transmitter and receiver) of the $j$-th ray in the $i$-th cluster, respectively. In particular, the number of clusters $N_{Ci}$ is commonly assumed to follow Poisson processes, while the arrival time difference between two clusters $T_{i} - T_{i-1}$ is assumed to be exponentially distributed. The angles of departure/arrival (AoDs/AoAs) in $a_{t,ij}$ and $a_{r,ij}$ can be assumed to be uniformly distributed within $[0, 2\pi]$, and the ray angle within a cluster can be assumed to follow zero-mean Laplacian distribution or a zero-mean second-order Gaussian mixture model (GMM). inquire Table III the $K$ factor in THz band is 12.9-21.6, which indicates that 90%-95% of path gain comes from the LoS cluster. Moreover, as the ASA and the ESA in THz are much smaller than their counterparts in cmWave and mmWave, we can assume only one ray in each cluster. Then, the THz channel can be modelled as a LoS rank-one matrix, i.e.,

$$H_{\text{LoS}} = \sqrt{G} \cdot a_i a_i^H,$$

(3)

where $G = \alpha G_t G_r$ is the channel gain. Due to the LoS dominant property, the beamforming technologies in the THz band show some unique aspects compared to those in the cmWave and mmWave band. First, the conventional beamforming is designed based on the estimated CSI, which accounts for the propagation on many (or even infinite) paths. While the THz beamforming merely takes into account of the LoS path, which can be realized without CSI by beam training (as will be detailed in Sec. IV). Second, LoS blockage is destructive for THz beamforming. Thus, the cooperative beamforming, e.g., IRS-assisted joint beamforming (as will be detailed in Sec. V), is worthy of consideration to reduce the LoS blockage.

2) Path Gain: The path gain of THz channels suffers from severe free spreading loss due to the extremely high frequency. According to the Friis’ formulation, the free spreading loss is given by

$$L_{\text{spread}}(f, d) = \left(\frac{4\pi fd}{c}\right)^2,$$

(4)

where $c$ is the speed of light in free space, $f$ is the carrier frequency, and $d$ is the path distance. Hence, the spreading loss increases with the frequency squared. In addition, the molecular absorption also causes severe attenuation of THz radial signals, which is not negligible. Hence, the path gain $\alpha_{ij}$ in (2) can be explicitly written as

$$\alpha_{ij}(f, d_{ij}) = \frac{1}{L_{\text{spread}}(f, d_{ij}) L_{\text{abs}}(f, d_{ij})} e^{-\frac{L_{\text{abs}}(f, d_{ij})}{\Gamma_{r}}} e^{-\frac{d_{ij}}{\Gamma_{e}}},$$

(5)

where $d_{ij}$ is the path distance of the $j$-th ray in the $i$-th cluster. $L_{\text{abs}}(f, d_{ij})$ is the molecular absorption which mainly comes from water vapor. $\Gamma_{r}$ and $\Gamma_{e}$ are the exponential attenuation factors of the arrive cluster and ray, respectively, which are frequency and wall-material dependent in general. Further information on these models is relegated to Table III. In general, the molecular absorption can be expressed as

$$L_{\text{abs}}(f, d_{ij}) = e^{k(f)d_{ij}},$$

(6)

where $k(f)$ is the total absorption coefficient that is comprised of a weighted sum of different molecular absorption in the medium, i.e., $k(f) = \sum m_i k_i(f)$, in which $m_i$ is the weight and $k_i(f)$ is the molecular absorption coefficient of the $i$-th species. The exact $k_i(f)$ on condition of any temperature and pressure can be obtained from the high resolution transmission (HITRAN) database. Based on the details therein, we plot $k(f)$ at frequencies from 1 to 10 THz in Fig. 3. As can be seen, the total absorption coefficient is relatively large at frequencies above 1 THz. Despite that two obvious drops can be witnessed in the range within 7-8 THz, the near-future THz systems are generally considered below 1 THz since the conventional metallic-antenna array requires extremely

---

**TABLE III**

| Evaluation Model | Description |
|------------------|-------------|
| MPM model [77]   | MPM is a broadband model for complex refractivity to predict the propagation effects of loss and delay for the neutral atmosphere. Input variables are barometric pressure, temperature, relative humidity, suspended water droplet concentration, and rainfall rate. |
| AM model [78]    | AM is a tool for radiative transfer computations at the microwave to millimeter wavelengths. The program can also be applied to a variety of other radiative transfer problems and has been used for applications ranging from laboratory receiver testing to radio spectrum management. |
| ITU-R model [79] | ITU-R P.676-11 provides methods to estimate the attenuation of atmospheric gases on terrestrial and slant paths by using a summation of individual gaseous absorption lines that is valid for the frequency range 1 GHz to 1 THz, based on atmospheric pressure, temperature, and water vapor. |

---

**TABLE IV**

| Total Molecular Absorption at some Frequencies |
|----------------------------------------------|
| $f$ (THz) | 0.14 | 0.26 | 0.35 | 0.41 | 0.67 | 0.85 |
| $k(f)$ (dB) | -42.2 | -38.5 | -27.8 | -22.4 | -18.5 | -20.9 |

---
describes how much power is transmitted/received compared to
provided model is accurate and reliable.

The striking jump points in both figures stand at the same
share the same trend of the loss on different transmit distances.

As shown in Fig. 4, both figures almost meanwhile, we plot the counterpart by using the ITU-R model
compute the corresponding path loss based on (5) and (6). In the

Let us consider the total path loss of the channel in (2).
By using the molecular absorption coefficient shown in Fig. 3
based on (6), we can obtain the exact path gain of \( \alpha_{11}(f, d_{11}) \)
in (5). By this means, we select \( k(f) \) within 0.1-1 THz and
compute the corresponding path loss based on (5) and (6). In the
meanwhile, we plot the counterpart by using the ITU-R model
tool for comparison. As shown in Fig. 4, both figures almost
share the same trend of the loss on different transmit distances.
The striking jump points in both figures stand at the same
frequencies, albeit with the difference of the order of magnitude.
Except for these jump points, the path gain calculated by our
provided model is accurate and reliable.

3) Antenna Gain: The transmit/receive antenna gain is the
product of element gain and array gain. The element gain
describes how much power is transmitted/received compared to
an isotropic element. Let \( G_e(\varphi, \theta) \) denote the element gain for
a propagation path with the azimuth angle \( \varphi \) and the elevation
angle \( \theta \), we have

\[
G_e(\varphi, \theta) = \zeta F(\varphi, \theta)D,
\]

where

\[
D = \frac{4\pi}{\int_{\varphi=0}^{\pi} \int_{\theta=0}^{\pi} F(\varphi, \theta) \sin \theta d\theta d\varphi}
\]

is the directivity, \( \zeta \) is the antenna efficiency, and \( F(\varphi, \theta) \) is
the normalized radiation pattern. The radiation pattern generally
focus more energy in the direction perpendicular to the element,
as shown in Fig. 5. An exemplary \( F(\varphi, \theta) \) is given by

\[
F(\varphi, \theta) = \begin{cases} \cos \theta, & \varphi \in [0, 2\pi] \text{ and } \theta \in \left[0, \frac{\pi}{2}\right] \\ 0, & \text{otherwise} \end{cases}
\]

If the radiation pattern has only one major lobe, the directivity
in (8) can be empirically calculated as

\[
D \approx \frac{4\pi}{\theta_{3\text{d}B}^e \theta_{3\text{d}B}^r (\text{radians})} = \frac{41253}{\theta_{3\text{d}B}^e \theta_{3\text{d}B}^r (\text{degrees})},
\]

where \( \theta_{3\text{d}B}^e \) and \( \theta_{3\text{d}B}^r \) is the half-power beamwidth (HPBW) of
the major lobe on x- and y-axis, respectively.

The array gain achieved if the signal is coherently added from
\( N \) array elements. Let \( P \) represent the transmit power. By using
\( N \) antenna elements (each equally assigned a power of \( P/N \)),
the maximum amplitude of the signal in the desired direction is
\( N \times \sqrt{P/N} = \sqrt{PN} \), and the corresponding signal power is
\( P N \). Thus, compared to using only one antenna element with
power \( P \), using \( N \) antenna elements achieves an array gain of
\( N \). As such, the antenna gain \( G_{t,ij} \) and \( G_{r,ij} \) in (2) are
respectively given by

\[
G_{t,ij} = N_e G_e(\varphi_{ij}, \theta_{ij}), \quad G_{r,ij} = N_e G_e(\varphi_{ij}, \varphi_{ij}).
\]

4) Noise: The noise term in (1) generally comes from two parts,
i.e., the electronic thermal noise and the re-radiation noise. The first part, also called background noise, is caused
by the thermal motion of molecules, which is related to the
frequency but independent of the transmitted signal power. Its
Considering the isotropic antenna elements, i.e., $G_e = 1$, we have $G_t = G_r = N$. The footprint of a uniform square array is approximately given by

$$S = \left(\frac{N \lambda}{2}\right)^2 \Rightarrow N = \frac{2f}{c \sqrt{S}},$$

where $\lambda$ is wavelength. For multi-input single-output (MISO) systems, there is no antenna gain at the receiver, and thus $G_r = 1$. In this case, based on (4) and (6), the received power can be rewritten as

$$P_r = P_t \cdot \frac{1}{f} \cdot e^{-k(f)d} \cdot \frac{c\sqrt{S}}{8(\pi d)^4},$$

where $1/f$ is the remaining component of the spreading loss. That is, the increased antenna gain cannot compensate for the spreading loss in MISO systems, and the received power decreases with the large-scale increase in frequency (without considering the small-scale fluctuation). In MIMO systems, the received power can be written as

$$P_r = P_t \cdot \frac{1}{f} \cdot e^{-k(f)d} \cdot \frac{S}{4(\pi d)^2},$$

where the antenna gain can compensate for the spreading loss and only the loss of molecular absorption remains. As can be seen from Fig. 3 the absorption factor does not show monotonicity with frequency. Thus, unlike MISO systems, the received power in MIMO systems might be higher with the large-scale increase in frequency, e.g., $P_r$ at 7.2 THz is larger than $P_r$ at 3 THz.

### B. Array Response Vector

Each array response vector $a_{r,ij}$ in (2) presents a propagation path on a specific AoD or AoA, which could be one or two dimensional, depending on the antenna geometry. Commonly, there are four typical array geometries: uniform linear array (ULA), uniform rectangular planar array (URPA), uniform hexagonal planar array (UHPA), and uniform circular planar array (UCPA). Next, we present the expression of the array response vectors in the above geometries and the corresponding path angles are plotted in Fig. 6.

1) **ULA (see Fig. 6(a)):** Considering an $N_a$-element ULA. The array response vector can be written as

$$a_{r,ij}(\phi)_{ULA} = \frac{1}{\sqrt{N_a}} \left[1, e^{jk(d_x \sin(\phi))}, \ldots, e^{jk(N_a-1)d_x \sin(\phi)}\right]^T,$$

where $d_x$ is the inter-element spacing, $k = \frac{2\pi}{\lambda}$ is the phase constant.

2) **URPA (see Fig. 6(b)):** Consider an URPA with $N_y$ times $N_z$ elements lying on the $yz$-plane. The array response vector is given by

$$a_{r,ij}(\phi, \theta)_{URPA} = \frac{1}{\sqrt{N_a}} \left[1, e^{jk(md_y \sin(\theta)) + nd_z \cos(\theta)}, \ldots, e^{jk((N_y-1)d_y \sin(\theta)) + (N_z-1)d_z \cos(\theta)}\right]^T,$$

where $d_y$ and $d_z$ are the inter-element spacing on the $y$-axis and $z$-axis, $0 \leq m \leq N_y - 1$ and $0 \leq n \leq N_z - 1$ are the

---

**Fig. 5. Normalized radiation pattern of antenna element.**
indices of antenna element, $N_a = N_y N_z$ is the total number of elements, and \( \varphi \) and \( \theta \) are the azimuth and elevation angles of arrival, respectively.

3) UHPA (see Fig. 6 (c)): Consider a UHPA consisting of \( V \) hexagon rings on the \( xy \)-plane. The inter-element spacing on the horizontal and vertical direction are \( d_x \) and \( d_y = \sqrt{3} d_x / 2 \), respectively, and the array response vector is given by

\[
a(\varphi, \theta)_{UHPA} = \frac{1}{\sqrt{N_a}} [f_V, \ldots, f_1, f_0, f_{-1}, \ldots, f_{-V}]^T, \tag{21}
\]

where \( \{f_v\}_{V=1}^{V} \) denote the ULA vectors at different rows, \( N_a = 1 + \sum_{v=1}^{V} v \), is the total number of elements. For different parities of the subscript \( v \), \( f_v \) can be expressed as

\[
f_v |_{v=2n} = e^{-jkd_a \sin(\theta) \cos(\varphi)} [ e^{-j(V/2-kd_x \sin(\theta) \cos(\varphi)}, \ldots, e^{-jkd_x \sin(\theta) \cos(\varphi)}, 1, e^{jkd_x \sin(\theta) \cos(\varphi)}, \ldots, e^{j(V-kd_x \sin(\theta) \cos(\varphi))}], \tag{22}
\]

\[
f_v |_{v=2n-1} = e^{-jkd_a \sin(\theta) \cos(\varphi)} [ e^{-j(V/2-kd_x \sin(\theta) \cos(\varphi))}, \ldots, e^{-jkd_x \sin(\theta) \cos(\varphi)}, e^{jkd_x \sin(\theta) \cos(\varphi)}, \ldots, e^{j(V-kd_x \sin(\theta) \cos(\varphi))}], \tag{23}
\]

4) UCPA (see Fig. 6 (d)): Consider a UCPA consisting of \( C \) circles on the \( xy \)-plane, each element uniformly distributed over the circle. The array response vector is given by

\[
a(\varphi, \theta)_{UCPA} = \frac{1}{\sqrt{N_a}} [1, \ldots, e^{jkr_c \sin(\theta) \cos(\varphi - \varphi_{nC})}, \ldots, e^{jkr_c \sin(\theta) \cos(\varphi - \varphi_{nC})}]^T, \tag{24}
\]

where \( N_a = 1 + \sum_{c=1}^{C} 6c \) is the total number of elements, \( r_c \) is the radius of the \( c \)-th circle, \( \varphi_{nC} \) is the angle of the \( n \)-th element in the \( c \)-th circle to the \( x \)-axis.

It should be mentioned that the array response vector can be set as a beamforming codeword to transmit a beam that covers a dedicated zone, as shown in Fig. 7(a). With increasing \( N_a \), the beam gain is improved, while the beam becomes narrower and thus more beams are needed to cover the whole space, which incurs difficulty in beam alignment and management. For example, Fig. 7(b) shows the UCPA beam patterns at \( \theta = 0 \), with \( N_a = 16 (4 \times 4) \) in MIMO systems and \( N_a = 1600 (40 \times 40) \) in UM-MIMO systems. The maximum beam gain raises from 16 to 1600, while the number of required beams raises from 64 (8 \times 8) to 6400 (80 \times 80). Table 4 shows the link budget for some THz scenarios with different antenna geometry.

C. Transceiver Architectures

In this subsection, we review the progress of transceiver architecture in THz UM-MIMO systems. The earliest beamforming architecture can be traced back to the attempt to construct high directional gain beams using phase shifters in antenna arrays, which is now known as analog beamforming. In analog beamforming architecture, all antennas adjust the phase of the same symbol through the phase shifters in the analog domain. Since the phase shifters can not change the magnitude of the symbol, the analog beamforming vector is subject to...
(DoF) [103]. Despite the ease of beam control, it is unaffordable to apply in UM-MIMO systems due to the high power consumption and high system cost [104].

As a cost-performance trade-off, hybrid beamforming (HB) architecture has emerged as an attractive solution for UM-MIMO systems [105], [106]. The HB can be expressed as a combination of analog beamforming and digital precoder, as presented in Fig. 8. HB architecture operates in both the baseband and analog domains, which has been shown to achieve the performance of the digital beamforming in some special cases but with much lower hardware cost and power consumption [107], [108]. There are mainly three types of HB architectures that have been reported extensively: the fully-connected architecture, the partially-connected (or array-of-subarray (AoSA)), and the dynamically-connected architecture.

1) Fully-Connected Architecture: As shown in Fig. 9 (a), the fully-connected architecture equipped with \( N_{RF} \) RF chains and \( N_t \) antenna elements. The output of each antenna element is the overlapped signals from all RF chains. The beamforming process of fully-connected HB architecture can be expressed as

\[
\mathbf{x} = \mathbf{F}_{AB} \mathbf{F}_{DP} \mathbf{s},
\]

where \( \mathbf{x} \in \mathbb{C}^{N_t} \), \( \mathbf{F}_{AB} \in \mathbb{C}^{N_t \times N_{RF}} \), \( \mathbf{F}_{DP} \in \mathbb{C}^{N_{RF} \times N_s} \), and \( \mathbf{s} \in \mathbb{C}^{N_s} \) denote the signal emitted by the antenna arrays, analog beamformer, digital precoder, and the transmitted symbol vector, respectively, and \( N_s \) is the number of data streams. The elements of each column in matrix \( \mathbf{F}_{AB} \) are phase shifters connected by one RF chain, while the elements in each row are phase shifters connected to one antenna port. In the fully-connected architecture, the antenna elements are fully used for every RF chain to provide high beamforming gain with a high complexity of \( N_t \times N_{RF} \) RF links (mixer, power amplifier, phase shifter, etc.) [109].

2) Partially-Connected Architecture: The partially-connected architecture is also known as AoSA. As shown in Fig. 9 (b), each RF chain is connected to a unique subset of antenna elements [110], i.e., a subarray. With this architecture, \( \mathbf{F}_{AB} \) in (25) has the form of a block diagonal matrix as \( \mathbf{F}_{AB} = \text{diag} \{ \mathbf{f}_1, \ldots, \mathbf{f}_{N_{RF}} \} \), where \( \{ \mathbf{f}_i \}_{i=1}^{N_{RF}} \) are \( \frac{N_t}{N_{RF}} \times 1 \) phase shifter vectors connected to the RF chains. \( \mathbf{F}_{AB} \) can also be

\[\begin{array}{|c|c|c|}
\hline
\text{Cases} & \text{Kiosk download} & \text{Data center} \\
\hline
\text{Frequency (THz)} & 0.14 & 0.22 \\
\text{Bandwidth (GHz)} & 5.00 & 20.00 \\
\text{Distance (m)} & 10.00 & 50.00 \\
\text{Propagation loss (dB)} & -95.36 & -113.27 \\
\text{Molecules absorption (dB)} & -0.0026 & -0.024 \\
\text{Thermal PSD (dBm/Hz)} & -174 & -174 \\
\text{Thermal noise (dBm)} & -77.01 & -70.99 \\
\text{Molecular noise (dBm)} & -92.36 & -103.27 \\
\text{Atmospheric attenuation (dB)} & 0.00 & 0.00 \\
\text{Tx antenna size} & 16 \times 16 & 32 \times 32 \\
\text{Rx antenna size} & 4 \times 4 & 32 \times 32 \\
\text{Rx antenna gain (dBi)} & 12.00 & 30.00 \\
\text{Noise figure (dB)} & 8.00 & 8.00 \\
\text{Other losses (dB)} & 10.00 & 10.00 \\
\text{SNR (dB)} & 2.52 & 9.70 \\
\text{Data rate (Gbps)} & 7.39 & 67.36 \\
\hline
\end{array}\]
Fig. 9. Three types of hybrid beamforming architecture: (a) fully-connected (FC); (b) partially-connected (PC); (c) dynamically-connected (DC).

### Table VI
Advantages and Limitations of Different Beamforming Architectures in the THz UM-MIMO.

| Beamforming Architectures | Advantages | Limitations | Feasibility to THz UM-MIMO |
|---------------------------|------------|-------------|---------------------------|
| Digital                   | Optimal performance, extremely high beam flexibility, full DoF. | Lots of RF chains, extremely high hardware costs, high power consumption. | Highly feasible but high overhead. |
| Analog                    | The simplest circuits, the lowest power cost, easy to implement. | Magnitude constraint or quantized bits, only support single stream, poor beam flexibility. | Highly feasible but limited DoF. |
| HB                        | FC         | Optimal or near-optimal performance with fewer RF chains, low power consumption. | Lots of phase shifters, difficult layout, difficult integration. | Medially feasible. |
|                           | PC         | Fewer phase shifters and fewer circuit lines compared to FC. | Each stream cannot obtain the array gain from all antennas. | Highly feasible but low array gain. |
|                           | DC         | Higher beam flexibility than PC. | Too many switches, difficult layout, difficult integration. | Medially feasible. |

expressed as

\[ F_{AB} = \Phi_{AB} M \]

where \( M \) can be regarded as a switch matrix of dimension \( N_t \times N_{RF} \), \( m_i \) is an all-one column vector, which means the state of corresponding phase shifters are on. The diagonal matrix \( \Phi_{AB} \) represents the phase shifters connected to the antenna elements, where \( \phi_i = e^{j\theta_i} \) is the phase factor of the \( i \)-th phase shifter. Compared with the fully-connected architecture, the partially-connected architecture further reduces the hardware cost to \( N_t \) links [109], [111], which is much appealing in THz communications [112], [113]. In particular, the authors in [114] proposed a novel partially-connected HB architecture with two digital beamformers, wherein the additional one is developed to compensate for the performance loss caused by frequency selective fading.

3) Dynamically-Connected Architecture: One disadvantage of the partially-connected architectures is the fixed circuit connection, which prevents adaptive and dynamic control [115]. To improve the beam flexibility, a dynamically-connected HB architecture was reported in [115]–[118]. As shown in Fig. 9 (c), a switching network \( W_S \) is added between the RF chains and the antenna elements [115], and the transmit signal can be written as

\[ x = \Phi_{AB} W_S F_{DP} s, \]

where \( \Phi_{AB} \) follows the definition in (26), \( W_S = \{ w_{i,j} \}, i = 1, \ldots, N_t, j = 1, \ldots, N_{RF} \), is a Boolean matrix, and \( w_{i,j} \) represents the switch from the \( i \)-th antenna element to the \( j \)-th RF chain. It is worth noting that since each antenna element is only allowed to connect to one RF signal at a time, \( W_S \) should satisfy constraint \( \sum_j w_{i,j} = 1 \). While maintaining the low-cost advantage of partial connection architecture, dynamic connection architecture improves processing freedom through the flexible switch network, which can be regarded as the
transition architecture from full connection to partial connection. In particular, the authors in [119] proposed a dynamic AoSA (DAoSA), where each antenna is connected to $N_{RF}$ phase shifters, claiming to save power compared to the fully-connected HB architecture. From the perspective of hardware structure, the DAoSA is more sophisticated than the fully-connected HB architecture, which may not be feasible for UM-MIMO systems. In addition, if we expect to balance the energy cost and performance between fully connected architecture and AoSA, a straightforward way is to turn off some phase shifters, rather than adding switches (which may cost additional power). We summarize the advantages and limitations of above architectures for THz UM-MIMO in Table VII.

### III. BEAMFORMING PRINCIPLES

In this section, we introduce the beamforming principles for THz UM-MIMO systems. To begin with, we specify some terms that are usually used in beamforming literature. Next, we endeavor to visualize how to steer a beam to desired directions via multiple antenna elements and unveil some important ideas behind it. Finally, we discuss the far-field and near-field assumptions in THz UM-MIMO systems.

### A. Nomenclature

Considering a point-to-point system in [1], the received signal after spatial post-processing can be expressed as $\mathbf{y} = \mathbf{W}^H \mathbf{H} \mathbf{s} + \mathbf{W}^H \mathbf{n}$, in which $\mathbf{F} \in \mathbb{C}^{N_t \times N_s}$ is the beamformer and $\mathbf{W} \in \mathbb{C}^{N_r \times N_s}$ is the combiner. The processing from the data stream $\mathbf{s}$ to the transmit signals $\mathbf{x} = \mathbf{F} \mathbf{s}$ is called beamforming. On the contrary, the processing from the received signals $\mathbf{y}$ to the data stream $\hat{\mathbf{y}} = \mathbf{W}^H \mathbf{y}$ is called combining. This process can be applied in the digital domain (e.g., by FPGA), or analog domain (e.g., by phase shifters), or in both domains. Thus, the beamforming technologies can be specified as digital beamforming, analog beamforming, and hybrid beamforming.

Rigorously, precoding is referred to as the digital case of beamforming, wherein the combining of this case is named decoding. The processor for precoding and combining is called precoder and decoder, respectively. For ease of presentation, most papers do not strictly distinguish the terms beamforming and precoding in the digital domain, i.e., digital beamforming is equivalent to precoding. beam steering is considered as a special case of analog beamforming with $N_s = 1$. In particular, in beam steering schemes, each beam is realized by an array response vector and the signal energy is steered to a specific direction [127]. The features of precoding/decoding, beamforming/combining, and beam steering are summarized in Table VII.

### B. Beam Steering

As a classic and basic technology to realize beamforming, the phased array has been well developed over decades. A phased array can steer the beam electronically in different directions, without moving the antennas [120]. Specifically, the power from the RF chain is fed to the antenna elements via a phase shifter on each. By this means, the radio waves from the separate antennas add together to increase the radiation in the desired direction, while suppress radiation in undesired directions. If phased arrays are employed at both transmitter and receiver, the resulting signal can be written as $y = w^H \mathbf{H} s + w^H \mathbf{n}$, where $\mathbf{f} \in \mathbb{C}^{N_t \times 1}$ is the beamformer with $N_t$ phase shifters, $\mathbf{w} \in \mathbb{C}^{N_r \times 1}$ is the combiner with $N_r$ phase shifters, and $s$ is a data symbol. The phase shifter only adjusts the phase without changing the amplitude of signal, which is subjected to a constraint, i.e., $|f(i)| = \sqrt{P}$ and $P$ is the total transmit power. Since there is only one RF chain, the beamforming via phased array is commonly known as analog beamforming.

Consider a LoS THz channel given in (8). It can be verified that the received signal vector has a form of array response vector, i.e., $y = H f s = a \cdot a_r$, where $a$ is a complex constant. In Section II-B, we straightforwardly present the expressions of $a_r$ for different types of arrays. Here, we focus on a simple example, i.e., ULA, to illustrate the relation between the mathematical expression [17] and its physical mechanism.

Fig. 10 plots the incoming signal wave to the receiver, in which $d_a$ is the antenna space and $\varphi$ is the arrival angle. It is obvious that for the same wavefront, the rightmost element receives it first, and the leftmost element receives it last. The wave-path difference between adjacent elements is $d_a \sin \varphi$. As

| Processor name | Beamforming/combining | Precoding/decoding | Beam steering |
|---------------|-----------------------|--------------------|--------------|
| Digital domain | Beamformer/combiner   | Precoder/decoder   | Beamformer/combiner |
| Analog domain | ✓                      | ✓                  | ×             |
| Data stream   | $N_s \geq 1$          | $N_s \geq 1$      | $N_s = 1$     |
| Codebook      | Not necessary         | Not necessary      | Necessary    |
| Applications  | MIMO and radar system | MIMO system        | Radar system  |
| Main advantages | Pattern diversity & Spatial multiplexing | Pattern diversity & Spatial multiplexing | Pattern diversity |
the distance increases $\lambda$, the phase increases $2\pi$. Thus, the phase difference between adjacent elements is

$$\frac{2\pi d_a \sin \varphi}{\lambda}. \tag{30}$$

As such, if the received signal at the first element is $c_r \in \mathbb{C}$, the received signal at the $n$-th element is $c_r \cdot e^{-j \frac{(n-1)2\pi d_a \sin \varphi}{\lambda}}$. As a result, the normalized array response vector for the receive ULA can be written as

$$a_r(\varphi) = \frac{1}{\sqrt{N_r}} \begin{bmatrix} 1, e^{-jkd_a \sin \varphi}, ..., e^{-jkd_a(N-1) \sin \varphi} \end{bmatrix}^T, \tag{31}$$

where $k = 2\pi/\lambda$. Next, we study how to use the phased array to combine the received signals. Substituting (29) into (28), we aim to find the optimal combiner to maximize the resulting power, which is equivalent to

$$P(1): \max_w |w^H a_r(\varphi)|^2$$

s.t. $|w(i)| = \frac{1}{\sqrt{N_r}}, \quad i = 1, 2, ..., N_r.$

It is easy to verify that an optimal solution is given by

$$w = a_r(\varphi), \tag{32}$$

The combining process can be regarded as an inverse beam-forming. As $a_r(\varphi)$ can maximally combine the signal from the direction of $\varphi$, we can also refer to this process as receiving a narrow beam in the direction of $\varphi$. \[121\]

The array response vector in (31) represents the signal coming from the direction $\varphi$. Vice versa, if the array transmits signal by using the array response vector in (31) as beamformer, the wavefront moves in the direction of $\varphi$. To be exact, as shown in Fig. 11, the signal waves in the direction of $\varphi$ have the same phase and will add together to increase the radiation power. Generally, the antenna space is considered to be half-wavelength, i.e., $d_a = \lambda/2$, to reduce the self-interference. For ease of expression, we assume that $N_t = N_r = N_a$, then the array response vector at both the transmitter and the receiver can be unified as

$$a(\varphi) = \frac{1}{\sqrt{N_a}} \begin{bmatrix} 1, -e^{j\pi \sin \varphi}, ..., -e^{-j\pi(N_a-1) \sin \varphi} \end{bmatrix}^T. \tag{33}$$

C. Far-field and Near-field

We should mention that all the array response vectors discussed above are based on an essential assumption, that is, the wavefront is the flat plane and all the elements have the same AoAs and AoDs. This assumption holds when the RF source is far away from the receiver. To illustrate this point intuitively, Fig. 12 plots the radiation cases of near-field and far-field. As can be seen, when the RF source is far away, the large radius of the spherical wavefront results in wave propagation paths that are approximately parallel. As such, we have $\theta_1 = \theta_2 = ... = \theta_M$. With the near RF source, the AoA varies for each element, i.e., $\theta_1 > \theta_2 > ... > \theta_M$. In other words, the far-field array response vector is inaccurate if applied to the near-field scenarios, or cannot maximally combine energy from a near-field source. Thus, an interesting question is when can we make the far-field assumption? In general, the boundary between near-field and far-field is given by the Rayleigh distance (RD) \[122\], i.e.,

$$RD = \frac{2D^2}{\lambda}, \tag{34}$$

where $D = (N_a - 1)d_a$ is the side length of the antenna array. That is to say, with the same aperture of antenna array, the RD for THz communication could be quite large due to its extremely small wavelength $\lambda$. As the THz frequency is thousands of orders higher than microwave frequency, it seems that the RD could be many kilometers. Is it true?

In fact, from another perspective, the THz antenna elements are commonly packed in a small footprint as the element spacing holds $d_a = \lambda/2$. As such, by substituting $D = (N_a - 1)\lambda/2$, the RD can be rewritten as

$$RD = D(N_a - 1) = (N_a - 1)^2\frac{\lambda}{2}. \tag{35}$$
which indicates that with the same number of antenna elements, the RD for THz communication could be quite small. Compared to the mmWave MIMO, THz UM-MIMO has more antennas but a smaller aperture of the array. Thus, there are two opposite effects on RD. Moreover, although the far-field array response model is inaccurate when the distance of the source is slightly less than RD, the array gain may not lose much if the far-field beam is still used. In sight of this, the authors in [123] derived the effective Rayleigh distance (ERD), to show the boundary at which the array gain loss of the far-field beam is 5%. The ERD is smaller than conventional RD and related to the AoA $\phi$, which is given by

$$\text{ERD} = 0.367\cos^2 \phi \frac{2D^2}{\lambda}. \quad (36)$$

Table VIII presents some representative RD and ERD in THz UM-MIMO systems, where the operation frequency is 0.3 THz. It can be seen that even using 10,000 antenna elements, the far-field beamforming is effective (with performance loss less than 5%) when the communication distance is more than 1.8 m. Thus, the far-field beamforming is still suitable for most scenarios while the near-field beamforming can be a supplement for a few scenarios.

Table VIII
| Array size | Antenna number | RD | ERD |
|------------|----------------|----|-----|
| 10 x 10    | 100            | 0.04 m | ≤ 0.15 m |
| 50 x 50    | 2500           | 1.20 m | ≤ 0.44 m |
| 100 x 100  | 10000          | 4.9 m  | ≤ 1.80 m |

In this section, we introduce a beam training scheme that can realize the beamforming in THz UM-MIMO systems without using the CSI. To start with, the basic concept of beam training is introduced. Then, two essential components in the beam training, i.e., training protocol and codebook design, are discussed, subsequently. Finally, we briefly introduce the lens antenna array to show a special manner to realize beam training.

A. Basic Concept

To enable a beam-connected wireless communication, the beamformer $f$ and the combiner $w$ need to be optimized to maximize the decoding SNR, which is equivalent to

$$\text{P}(2) : \{w^{\text{opt}}, f^{\text{opt}}\} = \arg \max \left| w^H H f \right|^2$$

s.t. $\|f\|^2 = 1$, $\|w\|^2 = 1$.

Provided that $H$ is known at the transmitter and receiver by channel estimation, the optimal beamformer $f^{\text{opt}}$ and the combiner $w^{\text{opt}}$ can be obtained by applying the singular value decomposition (SVD) on $H$. However, the conventional channel estimation methods may not apply to THz UM-MIMO systems due to the unaffordable operation complexity on large-scale array. Besides, the beam pattern of conventional pilot signals is almost omni-directional, which suffers from severe path loss over THz channels and may not be effectively detected by the receiver. Regarding the above issues, an interesting question is whether we can find $f^{\text{opt}}$ and $w^{\text{opt}}$ without the need of CSI?

The beam training scheme provides a promising solution [124]–[126]. Since the LoS path is dominant in THz channel, we can replace $H$ in (2) with the LoS channel in (3), i.e.,

$$\text{P}(3) : \{w^{\text{opt}}, f^{\text{opt}}\} = \arg \max \left| w^H a_r a_t^H f \right|^2,$$

where the optimal solution is given by the array response vectors, e.g., $w^{\text{opt}} = a_r$ and $f^{\text{opt}} = a_t$. Due to this fact, the feasible region of $f$ and $w$ can be reduced from a vector space to an AoA/AoD space. Thus, we can redefine a codebook $\mathcal{F}$ that includes many antenna response vectors representing the narrow beams corresponding to different AoAs/AoDs [127]–[129]. For ULA, the beamforming codebook can be written as

$$\mathcal{F} = \{a(\phi_1), a(\phi_2), ..., a(\phi_N)\}, \quad (37)$$

where $N$ is the number of beams. Despite $\phi \in [0, 2\pi)$, it suffices to consider the $N$ beams only within $[-\frac{\pi}{2}, \frac{\pi}{2}]$ (equivalent to $[0, \frac{\pi}{2}] \cup [\frac{3\pi}{2}, 2\pi)$) due to the following lemma [130].

**Lemma 1.** Beams within $[-\frac{\pi}{2}, \frac{\pi}{2}]$ and beams within $[\frac{\pi}{2}, \frac{3\pi}{2}]$ are isomorphic for ULA. In particular, the narrow beam in direction of $\phi$ is equivalent to that in direction of $\pi - \phi$, i.e.,

$$a(\phi) = a(\pi - \phi). \quad (38)$$

As can be seen from Fig. 13(a), each beam shows the symmetrical radiation patterns within the range $[-\frac{\pi}{2}, \frac{\pi}{2}]$ and $[\frac{\pi}{2}, \frac{3\pi}{2}]$, which validates Lemma 1. Moreover, the beam is narrower with the AoA/AoD around 0 and is wider with the AoA/AoD around $\pm \frac{\pi}{2}$. This implies that with the same
number of beams, the codebook with non-uniformly distributed beams may achieve a higher worst-case performance than that with uniformly distributed beams (see Fig. 13(b)). The optimal distribution of the beams’ AoA/AoD that maximize the worst-case performance is uniform distribution in sine space \([131]\).

The larger the codebook size, the closer the performance of beam training is to that of the optimal beamforming. Generally, the number of beams \(N\) required in the codebook is proportional to the number of array antennas \(N_a\). Fig. 14 shows the cases of \(N = N_a\) beams and \(N = 2N_a\) beams covering \([-\pi, \pi]\), where the AoA/AoD of the beams are uniformly distributed in sine space. The normalized worst-case performance of the above two cases is given by

\[
\rho = \begin{cases} 
\frac{1}{N_a \sin \left(\frac{\pi}{2N_a}\right)}, & N = N_a \\
\sqrt{2} \frac{N_a}{2N_a \sin \left(\frac{\pi}{4N_a}\right)}, & N = 2N_a
\end{cases}
\]

The above process is referred to as **exhaustive beam training**. Generally, the beam training strategy includes training protocol and codebook design. The protocol of the exhaustive beam training is the simplest, that is, exhaustively testing the narrow beam pairs of the transmitter and the receiver. The codebook consists of array response vectors, where the AoAs or AoDs are uniformly distributed in the sine space. While other beam training strategies requires more complicated training protocol and codebook design, which are discussed in the next subsections.

**Remark 1.** Sometimes, the concept of beam training is confused with beam alignment. Commonly, the beam alignment is to find a wide-beam pair to initialize a reliable connection, while the beam might be further optimized before transmitting data \([132]-[134]\). Whereas, beam training focuses on seeking the strongest narrow-beam pair and then directly using this beam pair for data transmission.

### B. Training Protocol

The training protocol determines the training reliability and complexity. Even though the exhaustive search protocol yields high reliability, it incurs the training complexity of \(N^2\) and thus is quite time-consuming in UM-MIMO systems. Compared to the exhaustive search, some other training protocols are more appealing owing to their lower training complexity.

For examples, IEEE 802.11ad proposed an one-side search protocol \([135]\), where each user exhaustively searches the beams in the codebook while the BS transmits the signal in an omnidirectional mode, which incurs the complexity of \(2N\). The authors in \([136]\) proposed a parallel search protocol which uses \(N_{RF}\) RF chains at BS to transmit multiple beams simultaneously while all users exhaustively search the beams, which incurs the complexity of \(N^2/N_{RF}\). The authors in \([137]\) proposed an adaptive search and the complexity is given by \(4 \log_2 N\). The authors in \([133]\) proposed a two-stage training scheme that combines sector level sweeping and fine search, which results in the complexity of \(N^2/Q + Q\), where \(Q\) is the number of narrow beams covered in each sector level. To the best of our knowledge, the most efficient protocol is **M-tree search** \([124]\), which is widely adopted in the hierarchical beam training.

In the M-tree search, we search the beams stage by stage with decreasing beam width. Specifically, there are \(S = \log_M N\) stages,
stages and the $s$-th stage has $M^s$ beams. Fig. 15 shows a feasible zone division of the ternary-tree codebook,
\footnote{It is worth mentioning that the zone division is a part of the protocol design and the uniform space division in Fig. 15 is not optimal.} i.e.,
\begin{equation}
M = 3 \text{, where } \omega^s_n \text{ denotes the } n \text{-th beam in the } s \text{-th stage.}
\end{equation}
As can be seen, each wide beam exactly covers three narrower beams in the next stage. Fig. 16 illustrates the diagram of the tree search. Specifically, we start with using an omnidirectional beam (root) for initial detection. Then, in each stage of the $M$-tree search, we find and follow the best beam (node) for the next stage search, until the best narrow beam (leaf) is found. It is worth mentioning that the $M$-tree search can be implemented on one side or both sides, as shown in Fig. 17. In the one-side tree search \footnote{We mainly discuss the approaches of using all antenna elements in fully digital beamforming. Other approaches such as antenna deactivation \footnote{We mainly discuss the approaches of using all antenna elements in fully digital beamforming. Other approaches such as antenna deactivation \cite{142}, changing frequency \cite{143}, and analog beamforming \cite{124} are out of the scope.}, we fix the transmitter to be in an omnidirectional mode and run an $M$-tree search stage by stage to find the best receive narrow beam. And then we fix the receiver to be in a directional mode with the found best receive narrow beam, and then run the $M$-tree search stage by stage to find the best transmit narrow beam. Thus, the complexity of one-side $M$-tree search is given by
\begin{equation}
T_{\text{one}} = M \log_M N + M \log_M N = 2M \log_M N.
\end{equation}
In the both-side tree search \footnote{We mainly discuss the approaches of using all antenna elements in fully digital beamforming. Other approaches such as antenna deactivation \cite{142}, changing frequency \cite{143}, and analog beamforming \cite{124} are out of the scope.} \footnote{We mainly discuss the approaches of using all antenna elements in fully digital beamforming. Other approaches such as antenna deactivation \cite{142}, changing frequency \cite{143}, and analog beamforming \cite{124} are out of the scope.}, we realize the beam training by selecting beam pairs stage by stage with decreasing beamwidth, where the receiver determines the best pair in each stage and feedback to the transmitter for the search (within the last selected range) in the next stage. Thus, the complexity of both-side $M$-tree search is given by
\begin{equation}
T_{\text{both}} = M^2 \log_M N.
\end{equation}
Notice from (40) and (41) that the complexity of the one-side $M$-tree search is less than that of the both-side $M$-tree search when $M \neq 2$, and the complexity is the same when $M = 2$. Besides, the both-side $M$-tree search needs $S$ feedbacks while the one-side $M$-tree search does not need any feedback. In particular, the ternary-tree search has the lowest searching complexity among all $M$-tree search schemes \footnote{We mainly discuss the approaches of using all antenna elements in fully digital beamforming. Other approaches such as antenna deactivation \cite{142}, changing frequency \cite{143}, and analog beamforming \cite{124} are out of the scope.}. Table IX summarizes the complexity of different training protocols.

![Fig. 16. Diagram of the ternary-tree search when $N = 27$.](image1)

![Fig. 17. Illustration of the one-side and both-side hierarchical search.](image2)

**TABLE IX**

| Training Protocol | Training complexity |
|-------------------|---------------------|
| Exhaustive search | $N^2$               |
| One-side search   | $2N$                |
| Adaptive search   | $4 \log_2 N$        |
| Parallel search   | $N^2/N_{RF}$        |
| Two-stage search  | $N^2/Q + Q$         |
| One-side $M$-Tree search | $2M \log_M N$ |
| Both-side $M$-Tree search | $M^2 \log_M N$ |

---

**C. Codebook Design**

Codebook design aims to optimize the weights of each codeword to make its beam pattern cover the zone specified by the protocol. For example, in the $M$-tree search, the narrow beams in the bottom stage can be set as array response vectors whereas the wide-beam design is an open problem.
The authors in [138] proposed an iterative coordinate decent where

\[
\hat{d}_{\lambda_1} \quad \text{as such, a complex target vector corrected as}
\]

\[
\text{degree of freedom in the objective vector. As we expect the}
\]

\[
\phi \quad \text{holds true for all } \varphi \quad \text{with}
\]

\[
\text{The COP is extended from ROP by adding a phase}
\]

\[
\text{2) COP: The COP is extended from ROP by adding a phase}
\]

\[
\text{As an approximate solution, one can minimize the Euclidean}
\]

\[
\text{P(4): min}_w \; \|AA^H w_{\theta^-} - \hat{d}_{\theta^-} \|.
\]

\[
\text{Then, the optimal solution to P(4) can be written as}
\]

\[
w_{\text{ROP}} = (AA^H)^{-1} A d_{\theta^-}.
\]

\[
\text{3) SNB: The two narrow beams } a_{N_1}(\varphi_1) \text{ and } a_{N_1}(\varphi_2) \text{ are orthogonal if the angles hold}
\]

\[
|\sin \varphi_1 - \sin \varphi_2| = \frac{2}{N_t}.
\]

\[
\text{For ULA with } N_t \text{ antennas, there are } N_t \text{ orthogonal narrow beams within } [-\frac{\pi}{2}, \frac{\pi}{2}], \text{ whose angles can be expressed as}
\]

\[
\varphi_n = \arcsin \left[ \frac{2n - 1}{N_t} \right], \quad n = 1, 2, ..., N_t.
\]

\[
\text{To construct a wide beam, an intuitive idea is by adding the}
\]

\[
\Psi = \{ \varphi_n \mid \theta^- \leq \varphi_n \leq \theta^+ \}.
\]

\[
\text{Then, the SNB solution can be written as}
\]

\[
w_{\text{SNB}}(\theta^-, \theta^+) = \sum_{\varphi_n \in \Psi} a_{N_1}(\varphi_n).
\]

\[
\text{4) SNB-GP: The major defect of the SA method is that}
\]

\[
P(6) : \min_{\{\omega_n\}_{n \in \Psi}} \text{var}[a_{N_1}(\theta^-) w_{\text{PSA}}(\theta^-)]
\]

\[
\text{s.t. } \theta^- \leq \phi \leq \theta^+,
\]

\[
\text{where } \text{var}[\cdot] \text{ represent the variance operator. P(6) is a NP-hard problem. To reduce the solving complexity, the authors in [121] show that}
\]

\[
\text{Finally, the variables } \{\omega_n\}_{n \in \Psi} \text{ in SNB-GP solution are obtained by exhaustive search on } \omega.
\]

\[
\text{5) SCA-ATP: The above approaches are heuristic and lack a mathematical metric. In sight of this, the authors in [140] proposed a beam-pattern error (BPE) metric to characterize the gap between the practical beam pattern and the ideal one. The BPE can be expressed as}
\]

\[
\varepsilon \{w_{\theta^-}, \theta^+) \} \triangleq \int_{\varphi \in [\theta^-, \theta^+]} \left| 1 - a_{N_1}(\varphi)^H w_{\theta^-} \right|^2 d\varphi
\]

\[
+ \int_{\varphi \in [\theta^-, \theta^-] \cup [\theta^+, \theta^+]} |\mu(\varphi) - a_{N_1}(\varphi)^H w_{\theta^-}|^2 d\varphi
\]

\[
+ \int_{\varphi \in [-\frac{\pi}{2}, \frac{\pi}{2}] \cup [-\frac{\pi}{2}, \frac{\pi}{2}]} |a_{N_1}(\varphi)^H w_{\theta^-}|^2 d\varphi
\]

\[
\text{The authors in [138] proposed an iterative coordinate decent method to solve P(5) and get the codeword } w_{\text{COP}}.
\]
Finally, the authors in [140] proposed an SCA-A TP algorithm to solve \( P(7) \) and get the codeword \( \mathbf{w}_{SCA-ATP} \).

6) S-SARV: The conventional array response vector assumes that the signal phase at the first antenna is zero. Here, we define the symmetrical array response vector as

\[
\hat{a}_N(\varphi) = \frac{1}{\sqrt{N}} \begin{bmatrix} e^{j\pi(-\frac{N-1}{2})\sin \varphi} \\ \vdots \\ e^{j\pi(n\sin \varphi)} \\ \cdots \\ e^{j\pi(\frac{N-1}{2}\sin \varphi)} \end{bmatrix}^T,
\]

by assuming the signal phase at the center of ULA is zero, where \( n = -\frac{(N-1)}{2}, -\frac{(N-1)}{2} + 1, -\frac{(N-1)}{2} + 2, \ldots, \frac{(N-1)}{2} - 1 \). Define \( M \) angles \( (M \geq 2N_a) \) as

\[
\psi_m = \arcsin \left( \sin \theta^+ - \frac{1}{N_t} + \frac{(m-1)(\sin \theta^+ - \sin \theta^- + \frac{2}{N})}{M-1} \right),
\]

for \( m = 1, 2, \ldots, M \). Then, the authors in [213] proposed to construct the wide beam by directly adding the symmetrical array response vectors with above angles, i.e.,

\[
\mathbf{w}_{S-SARV} = \beta \sum_{m=1}^{M} \hat{a}_{N_t}(\psi_m),
\]

where

\[
\beta = \frac{\sqrt{\mathbf{Z}} \cdot \text{norm} \left[ \sum_{m=1}^{M} \hat{a}_{N_t}(\psi_m) \right]}{\sqrt{N_t \left( \sin \theta^+ - \sin \theta^- + \frac{2}{N} \right)}}.
\]

To the best of our knowledge, the beam pattern of the SCA-ATP is closest to the ideal one, and the S-SARV is the best approach to balance the performance and computational complexity. The beam patterns of the two approaches are presented in Fig. 19. Table VII summarizes the performance and complexity of different wide-beam approaches.

### D. Lens Antenna Array

For implementing beam training with low cost, a concept of beamspace MIMO came up recently. This concept is particularly referred to as the beamforming realized by using the special hardware that makes the wireless communication system more like an optical one [144]. By modeling the optical lens as the approximate spatial Fourier transformer, beam training will be re-considered from the antenna space to the beam space that has much lower dimensions, to significantly reduce the number of required RF chains [145]. Fig. 20 shows the architecture of the training tailored hardware, in which each beam has a dedicated circuit and only the needed beams will be selected by limited RF chains via the selecting network. Compared to the conventional digital/analog hardware, the advantage of the training tailored hardware is that there is no need to control antenna weights. Instead, only a simple link selection is required, which makes the beamforming realized faster and more accurately.

One popular training tailored hardware is the lens antenna array, which is composed of an electromagnetic lens and some antenna elements located in the focal region of the lens [146]. In general, there are three main fabrication technologies for the electromagnetic lens [147]: 1) by conventional antennas array connected with transmission lines with variable lengths [148, 149]; 2) by dielectric materials with carefully designed...
and receive signals on the antenna elements, which have large magnitude and reduce the dimension from $N\times N$ to $N_{RF}\times N_{RF}$, where $N_{RF}$ represents the number of RF chains of the transmitter and receiver respectively. This is referred to as the antenna selection or beam selection problem. For ease of understanding, an illustration of the antenna selection in beamspace multi-user MIMO system is presented in Fig. 22. The less the number of users served, the less the number of antenna elements selected.

V. IRS-ASSISTED JOINT BEAMFORMING

Despite UM-MIMO are able to offer great beamforming gains, its energy efficiency decreases with the increase of the number of antennas. Besides, the LoS blockage problem significantly impairs the performance of THz beamforming. In this subsection, we introduce an emerging wireless technology, i.e., IRS, which can significantly improve the spectral efficiency with much-reduced energy consumption and reduce the impact of LoS blockage.

IRS is a metasurface consisting of a large set of tiny elements (i.e., controllable reflecting elements), each being able to pas-
sively reflect the incident wireless signal by adjusting its phase shifts. By judiciously optimizing the phase shifts at IRS, the reflected signals of different elements can be added/counteracted in intended/unintended directions [168]. Compared to the conventional MIMO systems whose performance is determined by their channels, the IRS-assisted UM-MIMO systems provide a programmable and controllable wireless environment [169]. Given this advantage, the achievable data rates in IRS-assisted UM-MIMO systems can be enhanced by jointly optimizing the precoder/decoder (i.e., beamformer/combiner) at the transmitter/receiver and the phase shifts at the IRS. The earliest researches on IRS-assisted joint beamforming were aimed at the precoder/decoder (i.e., beamformer/combiner) at the transmitter/receiver and the phase shifts at the IRS. The earliest researches on IRS-assisted joint beamforming were aimed at the MISO IRS-assisted system [170]–[172]. With different optimization targets, transmit power minimization [170], weighted sum-rate maximization [171], energy efficiency maximization [172], multicast rate maximization [173], and latency minimization [174] have been investigated.

A. IRS-Assisted UM-MIMO Systems

Consider a point-to-point IRS-assisted UM-MIMO communication system as depicted in Fig. 23 where the base station (BS), equipped with \( N_t \) antennas, transmits \( N_s \leq N_t \) data streams to a user, equipped with \( N_r \) antennas, with the help of an IRS equipped with \( N_{IRS} \) passive elements. In the communication, the BS sends its data message \( s \in \mathbb{C}^{N_t \times 1} \), via a precoder \( F \in \mathbb{C}^{N_t \times N_s} \) to the user and the IRS simultaneously. Let \( H \in \mathbb{C}^{N_r \times N_t} \), \( M \in \mathbb{C}^{N_{IRS} \times N_t} \), and \( N \in \mathbb{C}^{N_r \times N_{IRS}} \) denote the channels from the BS to the user, from the BS to the IRS, and from the IRS to the user, respectively. The received signal at the IRS is first phase-shifted by a diagonal reflection matrix \( \Theta = \text{diag}(\beta e^{j\theta_1}, \beta e^{j\theta_2}, \ldots, \beta e^{j\theta_{N_{IRS}}}) \in \mathbb{C}^{N_{IRS} \times N_{IRS}} \) and then reflected to the user, where \( j = \sqrt{-1} \) is the imaginary unit, \( \{\theta_i \in [0, 2\pi]\}_{i=1}^{N_{IRS}} \) are the shifted phases, and \( \beta \in [0, 1] \) denote the amplitude of each reflection coefficient. As such, the overall received signal is expressed as

\[
y = \sqrt{P} \frac{H_{LoS} F s + H_{IRS} M s}{N_s} + n, \tag{63}
\]

where \( P \) is the total transmitted power and \( \|F\|_F^2 = N_s \). In addition, \( n \sim \mathcal{CN}(0, \sigma_n^2 I_{N_r}) \) is zero-mean additive Gaussian noise. The aim is to maximize the spectral efficiency by jointly optimizing the precoding matrix \( F \) and the phase shifts \( \{\theta_i\}_{i=1}^{N_{IRS}} \), subject to the power constraint at the BS and the uni-modular constraints on the phase shifts. Let \( v = [e^{j\theta_1}, e^{j\theta_2}, \ldots, e^{j\theta_{N_{IRS}}}]^H \) denote the phase shifter vector at the IRS, i.e., \( \Theta = \beta \cdot \text{diag}(v^\dagger) \). Define the effective channel in IRS-assisted UM-MIMO systems as \( H_{eff} = N \Theta M + H_{LoS} \). Thus, the IRS-assisted joint beamforming problem can be formulated as

\[
P(8) : \max_{F,v} \log_2 \det \left( I_{N_s} + \frac{P}{\sigma_n^2 N_s} H_{eff} F F^H H_{eff}^H \right) \tag{64}
\]

s.t. \( H_{eff} = N \Theta M + H_{LoS} \),

\[
\|F\|_F^2 = N_s, \quad \Theta = \beta \cdot \text{diag}(v^\dagger),
\]

\[
|v(i)| = 1, \quad i = 1, 2, \cdots, N_{IRS},
\]

where \( v(i) \) denotes the \( i \)-th entry of \( v \). \( P(8) \) is a quite hard optimization problem as the non-convexity remains on both the objective function and the constraint imposed by IRS’s phase shifts \( v \). To solve this problem, the authors in [175] proposed a sum-path-gain maximization approach to reach a suboptimal solution. Then, the authors in [176] proposed an alternating optimization (AO)-based method to reach a high-performance near-optimal solution albeit compromised on the computational complexity. In sight of this, the authors in [177] proposed a manifold optimization (MO)-based algorithm to achieve a better performance-complexity tradeoff.

Hitherto, there have been many works focusing on the channel estimation solutions [178]–[182] and the beamforming optimization problems [183]–[189] for various IRS-assisted UM-MIMO scenarios. However, it is practically inefficient to combine the channel estimation and the beamforming designs in UM-MIMO THz systems due to the extremely high implementation complexity. Thus, it is of vital importance to consider the beam training strategies for IRS-assisted systems.

The main challenges of the beam training for IRS-assisted UM-MIMO systems are attributed to the passivity of IRS, i.e., unable to transmit and receive beams. Thus, the authors in [190] proposed to place the IRS relatively still to the BS and developed a fast beam training scheme by treating the BS and IRS as a whole. By skipping the signal processing, the authors in [191] proposed a machine learning empowered beam...
training framework for IRS-assisted UM-MIMO systems. Next, we introduce a cooperative beam training procedure for IRS-assisted systems proposed in [192] and [130], by using diagrams to illustrate its core idea.

### B. Cooperative Beam Training

Assume that the IRS is placed on the same horizontal level as the BS and the user, i.e., without loss of generality, we do not consider the elevation angle of IRS. As shown in Fig. 24, the IRS-assisted system consists of six path angles $\varphi_{B,H}, \varphi_{U,H}, \varphi_{B,M}, \varphi_{R,M}, \varphi_{R,N},$ and $\varphi_{U,N}$. The cooperative beam training aims to find the narrow beams at these angles. For ease of exposition, we first define the training modes for the active terminal (BS and user) and the passive terminal (IRS), respectively. As shown in Fig. 25, we use a solid red line to represent the transmit beams, a solid-blue broken line to represent the receive beams, a solid red arrow and a solid yellow line to represent the incoming signal and the reflected signal, respectively. It is worth mentioning that in the return mode, the codewords of IRS are functions of the AoAs, i.e., $\Theta_{\text{ret}}(\varphi_{k}^\text{in})$. If the AoA is $\varphi_{k}^\text{in}$, the codeword $\Theta_{\text{ret}}(\varphi_{k}^\text{in})$ ensures that the AoD of the reflected signals is in the back direction $\varphi_{k}^\text{in} + \pi$, i.e.,

$$\Theta_{\text{ret}}(\varphi_{k}^\text{in})a(\varphi_{k}^\text{in}) = a(\varphi_{k}^\text{in} + \pi).$$

(64)

The codewords of the directions mode are functions of AoAs and AoDs, i.e., $\Theta_{\text{dir}}(\varphi_{n}^\text{in}, \varphi_{p}^\text{out})$. If the AoA is $\varphi_{k}^\text{in}$, the codeword $\Theta_{\text{dir}}(\varphi_{k}^\text{in}, \varphi_{m}^\text{out})$ ensures that the AoD of the reflected signals is $\varphi_{m}^\text{out}$, i.e.,

$$\Theta_{\text{dir}}(\varphi_{k}^\text{in}, \varphi_{m}^\text{out})a(\varphi_{k}^\text{in}) = a(\varphi_{m}^\text{out}).$$

(65)

The design of the return-mode and direction-model codebooks are provided in [192]. In the next, we first present a primary protocol of the IRS-assisted joint beam training, albeit with some shortcomings, to draw some basic insights. We divide the overall protocol into three phases to achieve different groups of measurements as illustrated in Fig. 26.

**Phase 1:** Shut down the IRS. Fix the BS to be in an omni-beam transmitting mode and the user sweeps the beam to find the desired path direction. By switching the operation of the BS and the user, we obtain $\varphi_{U,H}$ and $\varphi_{B,H}$.

**Phase 2:** Keep the user silent and fix the BS to be concurrently in an omni-beam transmitting and receiving mode. Then, the IRS successively sweeps the codewords in return mode, i.e., $\Theta_{\text{ret}}(\varphi_{n}^\text{in})$, which are predefined with specific time slots, and is known to all terminals. The best codeword is informed to the BS by determining the time slot with the strongest receiving signal. By switching the operation of the BS and the user, we obtain $\varphi_{R,M}$ and $\varphi_{R,N}$.

**Phase 3:** With the obtained $\varphi_{R,M}$ and $\varphi_{R,N}$, we fix IRS to optimally bridge the BS-IRS-user link by direction mode, i.e., $\Theta_{\text{dir}}(\varphi_{R,M}, \varphi_{R,N})$. Fix the BS to be in an omni-beam transmitting mode and the user sweeps the beam to find the desired path direction. By switching the operation of the BS and the user, we obtain $\varphi_{U,N}$ and $\varphi_{B,M}$.

Based on the above three phases, we can find all the path angles in IRS-assisted systems [192]. However, this strategy suffers from the following main drawbacks.

- Omni-beam may not be effectively detected in THz communication.
Concurrently transmitting and receiving beams result in interference.

Sweeping the narrow beams incurs high complexity.

To circumvent these issues, we develop a more practical protocol, which combines the ternary-tree search (at BS and users) and the partial search (at IRS). Let $N$ denote the number of discretized AoAs/AoDs. There are $(3N - 2)/2$ codewords in the transit/receive codebook and $2N + 1$ codewords in IRSs codebook. Details of the codebook design is provided in [130]. The diagram of the training protocol is presented in Fig. 27 and the process is described as below.

**Phase 1:** We aim to obtain the optimal codeword for IRS. To achieve the beam alignment, we first test $3 \times 3$ wide beams in 9 successive intervals with BS in the transmitting mode and user $k$ in the receiving mode. In each interval, the IRS successively searches the codewords $\{\Theta_n\}_{n=1}^{2N+1}$. For the IRS, there is only one beam pair that covers both the BS-IRS link and the IRS-user link. During the interval when this beam pair (aligned case) is used, the user will detect an energy pulse in the time slot when IRS uses $\Theta_{\text{dir}}(\varphi_{R,M}, \varphi_{R,N})$. Thus, the user can utilize the pulse slots to identify this optimal codeword for IRS.

**Phase 2:** We turn off the IRS and aim to obtain $\varphi_{B,H}$ and $\varphi_{U,H}$ via the following three steps. In step 1, 9 wide-beam pairs are tested for alignment. The user compares the received energy in 9 intervals and determines the aligned pair with the maximum power. The aligned pair is labeled by recording the beams chosen at both sides. In step 2, the BS transmits the labeled wide beam and the user uses a ternary-tree search to obtain $\varphi_{U,H}$. In step 3, the user transmits the labeled narrow beam and BS uses a ternary-tree search to obtain $\varphi_{B,H}$.

**Phase 3:** We aim to obtain $\varphi_{B,M}$ and $\varphi_{U,N}$ through three steps similar to Phase 2. We turn on IRS with the obtained optimal codeword, i.e., $\Theta_{\text{dir}}(\varphi_{R,M}, \varphi_{R,N})$. Note that there exist two propagation paths from the BS to the user, i.e., BS-IRS-user path and BS-user path. The BS-user path has been estimated in Phase 2. To estimate the AoA and AoD of the reflecting paths, we can use the ternary-tree search with a small modification. Specifically, in each stage of step 2, the receiver removes the signal component of the BS-user path, by signal processing, when determining the best beam. In step 3, the user transmits the labeled narrow beam and BS uses a ternary-tree search while removing the signal component of the BS-user path. By this means, we can obtain $\varphi_{B,M}$ and $\varphi_{U,N}$.

By applying the above three phases, all the path angles can be found in IRS-assisted systems, which completes the THz IRS-assisted beam training. We would like to point out that the exhaustive search protocol has $N^2 + N^4$ tests in IRS-assisted systems, whereas this training protocol has only $18N + 18$.
the performance gains of the IRS-assisted schemes are significant. IRS-assisted beam training in the non-blockage scenario. These results also validate the effectiveness of THz the performance gained by IRS-assisted schemes is still not able. BS-IRS-user link provide propagation paths. One can see that THz non-blockage case, where both the BS-user link and the bound. Fig. 29 shows the performance of different schemes in beam training scheme is close to the IRS-assisted capacity without IRS. Besides, the performance of THz IRS-assisted compared to the schemes with random IRS (phase shifts) and 12log3N – 3 tests. Next, we compare the performance of the beam training with the capacity bound of the schemes with optimal IRS phase shifts, random IRS phase shifts, and without IRS. The capacity bound, i.e., the performance of the optimal beam forming, can be obtained by assuming perfect CSI. We consider a THz IRS-assisted UM-MIMO indoor scenario, in which a BS, a user, and an IRS are at the three vertices of a triangle with side length of 5 m. The numbers of BS/user antennas and IRS elements are all 128. The number of narrow beams in the bottom of the training codebook is 243. The operating frequency is set to 140 GHz with background noise power –80 dBm. Fig. 28 shows the performance of different schemes in THz LoS-blockage case, where H_{LoS} = 0. In the non-IRS-assisted scheme, we treat IRS as an indoor wall whose first-order ray attenuation is randomly set to between 5.8 dB and 19.3 dB compared to the LoS. It is observed that the performance gains of the IRS-assisted schemes are significant compared to the schemes with random IRS (phase shifts) and without IRS. Besides, the performance of THz IRS-assisted beam training scheme is close to the IRS-assisted capacity bound. Fig. 29 shows the performance of different schemes in THz non-blockage case, where both the BS-user link and the BS-IRS-user link provide propagation paths. One can see that the performance gained by IRS-assisted schemes is still notable. This is because the IRS can provide additional aperture gains via controllable reflection, so as to increase the received power at the user. These results also validate the effectiveness of THz IRS-assisted beam training in the non-blockage scenario.

VI. MULTI-USER THz UM-MIMO BEAMFORMING

In a single-user MIMO system, the baseband digital beamforming needs linear processing on all antenna elements at the transmitter or receiver, whereas in a multi-user MIMO system, it is generally assumed that no coordination among antenna elements at different users, and thus the single-user beamforming techniques, e.g., SVD operation, are no longer feasible in multi-user case. In light of this, there are two scenarios to be investigated in terms of multi-user beamforming, i.e., downlink beamforming, where the BS transmits independent data streams to multiples users simultaneously, and uplink beamforming, where a group of users transmit their to the BS simultaneously. The main interest for multi-user MIMO is to distinguish independent data streams in the spatial domain, which is known as space-division multiple access (SDMA).

A. Multi-User UM-MIMO Channels

The two scenarios come with two kinds of multi-user MIMO channels, respectively. The downlink channel is referred to as a broadcast (BC) channel, while the uplink channel is referred to as a multiple access (MA) channel. The sum-rate is a common performance metric for multi-user MIMO systems and there are some notable differences between BC and MA. In the former, the transmitted signal is a superposition of the data streams intended for all users, subject to the total transmit power constraint. By comparison, in the latter, the signal transmitted from each user is affected by other co-scheduled users, subject to individual power constraints. Uplink and downlink transmission policies depend on the accuracy of the channel state information at the transmitter (CSIT) and receiver (CSIR), and the optimal beamforming and power allocation between BC and MA channels have an implicit connection, known as duality. In a multi-user UM-MIMO system with hybrid beamforming architecture as shown in Fig. 30 the MIMO channel can be classified as an antenna-port channel (denoted by $H_{\text{ant}}$) and a digital-port channel (denoted by $H_{\text{dig}}$). The analog beamforming is the bridge between $H_{\text{ant}}$ and $H_{\text{dig}}$.

B. THz UM-MIMO Beamforming Strategy

In the THz UM-MIMO systems, the array at the transceivers has an extremely large number of antennas, which results in a large-dimensional matrix $H_{\text{ant}}$. Thus, the accurate channel estimation for $H_{\text{ant}}$ is time consuming and the multi-user precoding incurs high complexity. Here, we propose a cost-efficient solution which adopts the AoSA architecture and as space-division multiple access (SDMA).
combines the beam training (in the analog domain) and the precoding (in the digital domain).

Specifically, compared to digital beamforming, the AoSA can provide comparable performance in the low-rank channel using fewer RF chains. By performing the beam training in the analog domain, the digital-port channel is expected to have a high gain. Then, the multi-user channel estimation and precoding are merely applied in the digital ports with a much smaller overhead. The beam training can be regarded as a compression process from $H_{\text{ant}}$ to $H_{\text{dig}}$. By applying the beam training, the system can find the LoS path in the multi-user UM-MIMO channel without the need of high-complexity calculations. Although this strategy sacrifices some DoFs in $H_{\text{ant}}$, its performance may not lose much given the fact that $H_{\text{ant}}$ is generally LoS dominant \cite{131}. In particular, the LoS link can provide two DoFs for spatial multiplexing by using different antenna polarization modes \cite{198}.

C. Linear Precoding Algorithms

The goal of precoding in the digital domain is to convey independent data streams to different users. It was shown that Costa’s dirty-paper precoding can achieve the capacity region of the Gaussian BC multi-user MIMO channel \cite{199}. However, it comes at the expense of non-linear processing. In comparison, linear precoding algorithms are more practical and well-explored in these years, which can be classified into the following three major categories:

1) Orthogonal Space-Division Multiplexing (OSDM): The core idea of this class is to completely eliminate the inter-user interference by dividing the space into multiple orthogonal subspaces and allocating them to different users or data streams. This method was first used in multi-user MISO scenarios, with the name channel-inversion or zero-forcing (ZF) \cite{200}. In the multiple-user MIMO scenarios, block-diagonalization (BD) \cite{201} is used to create orthogonal sub-spaces by projecting each user’s channel into the null-space of other users’ channels. Based on ZF and BD, many extended precoding methods are studied to save the transmit power, reduce the computational complexity, or combat the effect of noise, e.g., iterative null space-directed SVD (Iterative Nu-SVD) \cite{202}, eigen-based ZF (EZF) \cite{203}, coordinated BD (CBD) \cite{204}, regularized BD (RBD) \cite{205}, QR-Regularized BD (QR-RBD) \cite{206}, complex lattice reduction RBD (CLR-RBD) \cite{207}.

2) Successive Interference Elimination (SIE): The core idea of this class is to design each user’s precoding matrix one by one and to make it lie in the null space of previous users’ channels. As such, in a system with $K$ users, user $k$ optimizes its precoding to compensate for the interference received from users $1, 2, \ldots, k-1$, and subject to the constraint that it does not interfere with any of those users. By leveraging the information of previous users’ precoding matrices, one can eliminate the interference successively. Two specific algorithms are presented in \cite{201} and \cite{208}, which are referred to as successive optimization (SO) and successive zero-forcing (SZF), respectively. However, these approaches generally have high-computational complexity since the solution depends on the design ordering of users. For a system of $K$ users, there are $K!$ sequentially optimized solutions, and are needs to find the best ordering to optimize the performance.

3) Minimizing the Product of Mean Squared Error Determinants (PDetMSE): The core idea of this class is to transform the classic sum-rate optimization into a tractable form, i.e., minimizing the PDetMSE. Specifically, by assuming that a minimum mean-squared-error (MMSE) decoder is applied at the receiver, the rate maximization problem with a sum power constraint is equivalent to a PDetMSE minimization problem under a sum power constraint, which can be solved by sequential quadratic programming (SQP) \cite{197}. Based on this result, many approaches have been proposed to approximate or reformulate the objective function of PDetMSE, so as to simplify the difficulty. These approaches are known as minimizing the product of mean squared errors (PMSE) \cite{209}, minimum total mean square error criterion (T-MMSE) \cite{210}, and weighted sum minimum mean square error (WMMSE) \cite{211}.

The above algorithms have been proven to achieve decent performance under the Rayleigh channel, which models the antenna-port channel with small-scale arrays for frequency below 6 GHz. In THz UM-MIMO systems, the precoder is optimized based on the digital-port channel $H_{\text{dig}}$ rather than the Rayleigh channel, and thus the performance of these algorithms needs to be re-evaluated. Now let us consider a THz UM-MIMO system, where a BS serves 8 users randomly located 50 meters away on a 140 GHz carrier. The background noise power at the receiver is set to $-80$ dBm. The BS equips 16 RF chains each connected to a subarray with 256 antenna elements and each user equips 2 RF chains each connected to a subarray with 16 antenna elements. Assuming the beam alignment is successful in the analog domain with two antenna polarization modes, we select five low-complexity algorithms, i.e., BD \cite{201}, EZF \cite{203}, QR-RBD \cite{206}, PMSE \cite{209}, and T-MMSE \cite{210}, to seek solutions for the precoder to the digital-port channel. The averaged sum-rate over 1,000 random channel realizations
is shown in Fig. 31. It can be observed that EZF suffers notable performance loss and PMSE is no longer effective in a high-transmit power regime. BD with water-filling (WF) power allocation has a small improvement compared with BD between the power of 10 and 20dB. The performance of QR-RBD and T-MMSE are comparable, with the largest achievable sum-rate among the five algorithms. Thus, QR-RBD and T-MMSE might be the promising digital beamforming algorithms in THz multi-user UM-MIMO channels.

VII. WIDEBAND BEAMFORMING

Based on the IFFT technologies, digital wideband beamforming can be realized by concurrently applying precodings for channels of different frequencies. In this section, we focus on the wideband beamforming in the analog domain, i.e., using the phased array. There are two main effects, i.e., spatial-wideband effect and frequency-wideband effect, that need to be considered in the THz UM-MIMO systems. In sight of this, we introduce three solutions, true-time-delayer (TTD) beamformer, max-min beamformer, and beam split aggregation and multiplexing (BSAM) in the THz wideband communications.

A. Spatial-Wideband Effect

Now, we consider a wideband communication with a very high symbol rate. When the plane wave arrives slantingly in a large array, the signal envelope received by different antenna elements might not belong to the same symbol and thus the phased array cannot effectively combine the signals, which is called the spatial wideband effect [212]–[214].

Let $T_s$ represent the symbol period and the baseband signal can be expressed as $s(t) = \sum_i \text{sym}(i)g(t - iT_s)$, where $\text{sym}(i)$ is the $i$-th symbol and $g(t)$ is the pulse shaping function, i.e.,
\[
g(t) = \begin{cases} 
1, & 0 \leq t \leq T_s \\
0, & \text{otherwise}
\end{cases}.
\]

For simplicity, we consider a SIMO ULA system and assume that the equivalent baseband signal received by the first antenna, i.e., the rightmost element shown in Fig. 32 is $\alpha s(t)$, where $\alpha$ is the path loss. Since the time delay between adjacent elements is $\frac{d_a \sin \varphi}{c}$, the time delay at the $m$-th antenna element is
\[
\hat{\tau}_m = \frac{(m - 1)d_a \sin \varphi}{c}.
\]
The phase difference at the $m$-th antenna element is $2\pi \cdot \hat{\tau}_m \cdot f = \frac{2\pi(m - 1)d_a \sin \varphi}{\lambda}$. Thus, the equivalent baseband signal received by the $m$-th antenna element can be written as
\[
y_m(t) = \alpha s \left( t - \frac{(m - 1)d_a \sin \varphi}{c} \right) e^{-j\frac{2\pi(m - 1)d_a \sin \varphi}{\lambda}}.
\]

Let $\hat{\tau}_{\text{max}} = \frac{(N_t - 1)d_a \sin \varphi}{c}$ denote the maximum time delay. In the narrowband communication, $T_s$ is relatively large and we have $\hat{\tau}_{\text{max}} \ll T_s$. In this case, we can assume that $s \left( t - \frac{(m - 1)d_a \sin \varphi}{c} \right) = s(t)$. Thus, the received signals are simplified as $y(t) = \alpha s(\varphi)s(t)$. As such, the multipath narrowband channel can be modeled as $Q$. This indicates that the received signal on each element has merely the phase difference and phase compensation can be used to provide coherent combining at the receiver. However, in the wideband communication, i.e., $T_s$ is close to or even less than $\hat{\tau}_{\text{max}}$, the antenna’s first element and the last element may receive different symbols. Thus only using phase shifters for combination is infeasible. A larger number of elements $N_t$ yields possibly larger $\hat{\tau}_{\text{max}}$, leading to the spatial-wideband effect determined by the array size and symbol rate. To eliminate the spatial-wideband effect, we can adopt orthogonal frequency division multiplexing (OFDM) modulation and set its cyclic prefix duration larger than the maximum time delay.

B. Frequency-Wideband Effect

Even without the spatial-wideband effect, i.e., all antenna elements receive the same symbol, the signal components of different frequencies arrive at the array with different phase differences. Thus, using phased array that combines signals with the same phase compensation cannot simultaneously maximize the beam gains at all frequencies, which is called the frequency-wideband effect. In other words, the beam pattern changes with the frequency of the signal, which is known as the phenomenon of beam squint or beam split [215]–[217].

The beam squint effect takes place at both the transmitter and the receiver. Without loss of generality, we analyze its effect in terms of a receive ULA. Let $B$ and $f_c$ denote the baseband...
bandwidth and carrier frequency, respectively. The wideband response array vector can be expressed as

\[
\mathbf{a}(\psi, f) = \frac{1}{\sqrt{N_a}} \left[ 1, e^{-j \frac{2\pi f f_d a \sin \psi}{c}}, \ldots, e^{-j \frac{N_a - 12 \pi f f_d a \sin \psi}{c}} \right]^T,
\]

(69)

where \( f \in \left[ f_c - \frac{B}{2}, f_c + \frac{B}{2} \right] \). Commonly, the combiner is set based on the carrier frequency, i.e., \( \mathbf{a}(\varphi) \triangleq \mathbf{a}(\varphi, f_c) \). Thus, the normalized beam gain of the combiner \( \mathbf{a}(\varphi) \) in the direction of \( \psi \) at frequency \( f \) can be expressed as

\[
A_f(C(\varphi, \psi)) = \left| \sqrt{N_a} \mathbf{a}(\varphi)^T \mathbf{a}(\psi, f) \right|^2
\]

(70)

where \( \xi = \frac{f}{f_c} \). Consider a wideband incoming signals with AoA \( \psi = \pi/4 \), \( f_c = 0.14 \text{ THz} \), \( B = 10 \text{ GHz} \), \( N_a = 80 \). Fig. 33 plots the normalized beam gains, i.e., beam patterns, at frequency \( f_c \), \( f_c - \frac{B}{2} \) and \( f_c + \frac{B}{2} \). As we can see, compared to the beam pattern at \( f_c \), the other patterns have angle squints. It is worth mentioning that the beam squint direction is different between the transmitter and the receiver. To be exact, as shown in Fig. 33, the beam pattern received with high frequency moves towards 90 degrees while that with low-frequency moves towards 0. On the contrary, we would like to point out that the beam pattern transmitted with high frequency moves towards 0 while the low-frequency beam moves towards 90 degrees.

Let us revisit the normalized beam gain given in (70). For incoming signals with AoA of \( \psi \) at frequency \( f \), the optimal combiner \( \mathbf{a}(\varphi) \) satisfies

\[
\xi \sin \psi - \sin \varphi = 0
\]

\[
\Rightarrow \varphi = \arcsin \left( \frac{f}{f_c} \sin \psi \right).
\]

(71)

Thus, the beam squint angle is given by \( |\varphi - \psi| \). Here, we note some insights for the beam squint effects as below.

- When \( f = f_c \), there is no beam squint as \( \varphi = \psi \).
- Beam squint effect increases with \( \psi \).
- Beam squint effect increases with \( f_c \).

Due to the frequency-wideband effect, the signal at edge frequency suffers from performance loss even under a successful beam alignment. As a result, the minimum beam gain over \( [f_c - \frac{B}{2}, f_c + \frac{B}{2}] \) is given by

\[
A_f(\varphi) = \left| \frac{\sin \left( \frac{N_a \pi B}{4 f_c} \sin \varphi \right)}{\sqrt{N_a \sin \left( \frac{4 f_c}{B} \sin \varphi \right)}} \right|^2.
\]

(72)

Accordingly, the beam squint loss can be written as \( 1 - A_f(\varphi_{\text{max}}) \), in which \( \varphi_{\text{max}} \) is the maximum beam deflection. When the beam patterns at the highest and the lowest frequencies are splitting, the minimum beam gain is zero.

### C. TTD Beamformer

The reason for both spatial-wideband and frequency-wideband effects is the time-delay spread over the antenna arrays, which cannot be eliminated by phased array. Thus, the TTD beamformer can be used to address both effects, as the time-delay spread can be completely offset by TTD components. To realize the TTD, we can prolong the propagation length or slow the wave velocity of the electromagnetic signals in the precoder/combiner. The former approach has been mainly adopted in existing works. Some emerging THz TTDs have been reported in recent years. There are four feasible TTD architectures in wideband beamforming, as presented in Fig. 34. In Fig. 34 (a) and (b), full-connected TTD (FC-TTD) architecture and partially-connected TTD (PC-TTD) architecture replace the phase shift matrix with a time delay matrix, respectively. Although these architectures can directly apply the existing beamforming algorithms, THz time delay consumes more power than the conventional phase shifter, which is not cost-efficient when a large number of TTDs are required. To remedy this deficiency, delayed phase precoding (DPP) structures are used to provide a low-cost wideband effect compensation scheme, as shown in Fig. 34 (c). The TTD layer inserted in the RF chains and phase shifters can support joint control of delay and phase to provide spatially aligned frequency-dependent beams over the entire bandwidth, thereby mitigating frequency-wideband effects while reducing hardware cost. Similarly, the authors in proposed to combine the TTD devices and the conventional phase shifts to jointly design wideband beamforming with reduced implementation cost compared to the fully TTD arrays. The infinite-precision TTD is impractical for hardware implementation. As a solution, a switching net-
work is designed to flexibly select the appropriate time delay from a set of fixed TTDs, as shown in Fig. 34 (d). The hardware complexity of dynamic-subarray FTTD (DS-FTTD) is lower than that of FC-TTD, PC-TTD, and DPP architectures, since there is no need to adjust the delay \[66, 220.\]

D. Max-Min Beamformer

The limitation of the phased array in wideband communication is that the phase shifts are flat over the signal band, which cannot cater to the wideband array response vector in (69) with \( f \in \left[ f_c - \frac{B}{2}, f_c + \frac{B}{2} \right] \). The conventional beamformer that maximally combines the signal power at \( f_c \) might cause severe beam squint loss. If we do not change the array architecture, i.e., still using the phased array, a max-min beamformer is proposed to reduce the frequency-wideband effect.

Specifically, in the wideband UM-MIMO systems, the THz channel is frequency related and can be modelled as

\[
H_{\text{LoS}}(f) = \sqrt{\alpha_c}(f)\mathbf{a}_t^H(f), \quad f \in \left[ f_c - \frac{B}{2}, f_c + \frac{B}{2} \right].
\]  

(73)

Thus, the performance metric in wideband communication systems should also be frequency-dependent and is denoted by \( U(f) \). For example, the spectral efficiency of UM-MIMO can be expressed as

\[
U(f) = \log_2 \det \left( \mathbf{I} + \frac{P}{\sigma_n^2} \mathbf{H}_{\text{LoS}}(f)\mathbf{w}\mathbf{w}^H \mathbf{H}_{\text{LoS}}^H(f) \right). \tag{74}
\]

The design of the max-min beamformer \( \mathbf{w} \) needs to cater to the metric over a wide band, which is formulated as

\[
P(9) : \max_{\mathbf{w}} \min_{f \in \left[ f_c - B/2, f_c + B/2 \right]} U(f).
\]

To the best of our knowledge, there has been no work considering the max-min beamformer design in THz UM-MIMO yet, however, it is an effective solution for wideband communication and worthy of in-depth research in the future. To validate its effectiveness, we consider a beam gain metric and the problem is formulated as

\[
P(10) : \max_{\mathbf{w}} \min_{f \in \left[ f_c - B/2, f_c + B/2 \right]} \left| \mathbf{w}^H \mathbf{a}(\psi, f) \right|^2
\]

\[\text{s.t. } \mathbf{w}(i) = 1, \quad i = 1, 2, \ldots, N_a.\]

Let us consider a system with \( f_c = 100 \text{ GHz} \) and \( B = 10 \text{ GHz} \). By applying a proper algorithm to solve \( P(10) \) (for which the details are omitted), we design a max-min beamformer \( \mathbf{w} \) with AoA \( \psi = 40^\circ \) and its beam pattern and beam gain over the band is shown in Fig. 35 (left). It can be seen that the beams of the conventional beamformer are split and thus the wideband beam gain (i.e., the minimum beam gain over the band) is zero. In comparison, the max-min beamformer can form a wide beam, which avoids the beam misalignment under beam squint.

E. BSAM

An interesting concept of BSAM arises recently for realizing the wideband beamforming \[309\] with AoSA. The BSAM scheme is to use the conventional beamformer, i.e., array response vectors, on multiple phased arrays and judiciously optimize the frequency allocation for these arrays to simultaneously serve multiple users in wideband communications.

Specifically, if user \( k \) is not in the perpendicular direction of the array, the effective signal band at user \( k \) is limited due to the beam split effect. To satisfy the user’s wideband demand, a beam split aggregation scheme is proposed by using multiple subarrays to transmit multiple beams, respectively. As shown in Fig. 36 (left), each subarray transmits a beam with an independent subband, where subarray 1 uses beamformer \( \mathbf{w}_1 = \mathbf{a}(\phi_k, f_{c,1}^k) \) and subarray 2 uses beamformer \( \mathbf{w}_2 = \mathbf{a}(\phi_k, f_{c,2}^k) \), and the user aggregates the subbands to meet the wideband demand. Meanwhile, a beam split multiplexing scheme is proposed where each subarray may serve multiple users as shown in Fig. 36 (left bottom). To be exact, if we use \( \mathbf{a}(\phi_k, f_{c,1}^k) \) to transmit the beam with central frequency \( f_{c,1}^k \) to the user \( k \), we define a beam split coverage (BSC) for this beamformer as

\[
BSC = [\psi_k^{\min}, \psi_k^{\max}], \tag{75}
\]

where if \( \phi_k \geq 0 \) we have

\[
\psi_k^{\min} = \arcsin(f_{c,1}^k \sin \phi_k), \quad \psi_k^{\max} = \arcsin(f_{c,1}^k \sin \phi_k), \tag{76}
\]

in which \( f_{\min} \) and \( f_{\max} \) are the minimum and maximum frequency available in the systems, respectively. If \( \phi_k < 0 \), we need to switch the values of \( \psi_k^{\min} \) and \( \psi_k^{\max} \). When the \( j \)-th user is in the BSC, i.e., \( \phi_j \in \text{BSC} \), we can allocate a frequency band with bandwidth \( B_j^1 \) to user \( j \), and the band range is given by

\[
\Omega_j^1 = \left[ f_{c,1}^k \sin \phi_k - \frac{B_j^1}{2}, f_{c,1}^k \sin \phi_k + \frac{B_j^1}{2} \right]. \tag{77}
\]
where the signals within the band $\Omega_1$ will be steered to user $j$ by the common beamformer.

The beam split aggregation scheme can satisfy the user’s wideband demand but incurs a high cost since each subarray merely serves one user. On the other hand, the beam split multiplexing scheme enables one subarray to serve multiple users, but the effective bandwidth received by each user is limited. By combining their advantages, as shown in Fig. 36 (right), the BSAM scheme can satisfy the user’s wideband demand while reducing the number of subarrays used.

VIII. EXISTING THz MIMO ARRAYS

Currently, different fabrication techniques are considered to implement THz band antennas. These techniques can be roughly divided into three categories: electronic-based, photonics-based, and new materials-based. In the electronic approach, a variety of antenna types based on different materials and processes have been reported, including but not limited to horns [227], [228], reflectors [229], [230], and cavity-backed slot antenna arrays [231]. At the same time, methods such as photo-conductive antennas [232], [233] and silicon-based lenses [244]–[246] have been tried in the field of photonics. In addition, the latest option for developing antennas in the THz band is based on the new phase-change materials such as vanadium dioxide (VO2) [237], graphene [238] and liquid crystal (LC) [239]. A detailed investigation of graphene THz crystal is provided in [240].

Researches on individual THz band antennas reveal the possibility of making up the THz regions in the electromagnetic spectrum. However, the performance of individual THz antennas is mostly limited by the low transmit power or poor dynamic beam scanning capability. To this end, it is necessary to study the large-scale THz antenna array that supports high transmit power and adjustable direction. Fortunately, the very short wavelength of the THz band supports the integration of a large number of antenna arrays with a small footprint, which helps to achieve dynamic directional high-gain beamforming via antenna arrays. Although some methods have been tried for the implementation of THz antenna arrays [241]–[243], these arrays only enhance the directional gain and do not support dynamic beamforming. In the following, we will focus on array fabrication techniques that have the potential to support dynamic beamforming.

A. Electronic and Photonic Approaches

For supporting the integration of small-size on-chip antennas, silicon-based THz antennas, which are mainly supported by CMOS and silicon-germanium (SiGe) technologies, have been developing rapidly in recent years. Silicon-based technologies have the advantages of simple structure, easy array integration, small size, and potential low-cost [20]. Its operating frequency is mostly limited to 0.5 THz [244], [245], which can provide solutions in the lower frequency band of the THz spectrum. Due to the relatively mature integration capabilities of silicon-based, many small-scale phased arrays based on various silicon-based processes have been reported [244]–[245]. An example is given in [254], which shows a $4 \times 4$ URPA using 130 nm SiGe bipolar-CMOS (BiCMOS) technology at 320 GHz. As shown in Fig. 37 (a), this array is made up of 16 elements and a fully-integrated 160 GHz phase-locked loop (PLL), consisting of charge pump (CP), phase/frequency detector (PFD), current-mode logic (CML), and injection-locking frequency divider (ILFD). The authors in [223] propose a 280 GHz $2 \times 2$ chip-scale dielectric resonator antenna array. As shown in Fig. 37 (b), this array incorporates a balun for an X-band input signal, a $1:4$ Wilkinson divider, and four x27 active multipliers chains to drive the elements.

Despite the various progress that has been witnessed and is still ongoing in the field of silicon-based arrays, the drastic power drop associated with this technology is a major bottleneck [20]. In the photonic approach, schemes for THz dynamic beam scanning are designed. For instance, frequency-scanning antennas can be used to control THz beam steering [224]. In [255], the proposed photoelectric phase shifter can control 300 GHz beam scanning within 50 degree. Recently, the optical
TTD phase shifters are also employed to offer stable time delay for wideband communications [224]. Fig. 37 (c) presents the schematic view of an optical TTD-based chip, wherein the input optical signals will be converted to 300 GHz frequency region by the InP photomixers and finally radiated by the 1 x 4 bow-tie antenna.

B. New Materials Approach

In addition to the electronic and photonic processes, new materials provide another way to achieve high-performance THz antennas [256–258]. Graphene, i.e., a two-dimensional form of graphite, has attracted the attention of the scientific community due to its unique electronic and optical properties. Compared with conventional electronic materials, graphene is highly tunable, so it can be used to implement devices that support dynamic beamforming [259, 260]. Fig. 37 (d) shows the working principle and design of the THz front end in [259].

Graphene antennas at the THz band with reconfigurable radiation patterns have been developed in [261–263]. In addition to individual graphene antennas, small-scale graphene antenna arrays have been tried in [264–266], but the beam scanning range has not been practically tested. Furthermore, [267] proposed a reconfigurable MIMO antenna system for THz communications. [22] envisions the use of graphene to implement UM-MIMO plasmonic nano-antenna arrays, which can implement a 1024 x 1024 UM-MIMO system at 1THz with arrays that occupy just 1 mm². Liquid crystal and graphene also envision benefits.

To sum up, Table XI shows the reported THz antenna arrays with dynamic beam scanning capability.

IX. Emerging Applications

In this section, we introduce some applications of beamforming technologies for THz communication. Specifically, we discuss six major scenarios also shown in Fig. 38 and their envisioned benefits.
A. Satellite Communications

In the past, free space optical communication has been extensively studied for realizing satellite networks since it allows high-bandwidth data transmissions, which are difficult to be detected and intercepted [277]. Recently, research progress has shown the prospect of using THz communication to build satellite networks, thanks to its high bandwidth and negligible molecular absorption in this scenario. In addition to the advantages of high transmission rate and security of directional optical communication, THz communication has higher energy efficiency and easier beam control. In the future, the miniaturized THz communication systems can be applied to the high-speed inter-satellite communication of satellite clusters and satellite-to-ground communication to realize the integrated space-air-ground communication envisaged by 6G [278].

B. Vehicular Connectivity

To realize the concept of intelligent transportation system, vehicle communication network has been extensively studied, which roughly includes three types of scenarios: vehicle to vehicle (V2V), vehicle to infrastructure (V2I), and vehicle to anything (V2X) [45], [279]. Vehicle networks require high data rates, low latency, and reliable communications. For instance, real-time traffic demands a data rate of 50 Mbps, and auxiliary communication for autonomous driving requires a delay less than 10 ms and reliability of 99.999% [58]. For the era of wireless interconnected smart cars, THz communication is anticipated to provide ultra-capacity, low-latency, high-mobility, and high-reliability for safe autonomous driving and intelligent transportation system (ITS) [56], [280]. It has been shown that THz wireless communication can support high-speed mobile railway application scenarios through dynamic beam tracking technology [281]. The channel propagation characteristics of the 300 GHz carrier in the V2I scenario have been explored in [282], which can be used to support the link-level and system-level designs for future THz vehicular communications.

C. Wireless Local Area Networks

The capability of THz communication of achieving short-distance high-speed coverage in indoor scenarios can be leveraged to enhance the WLAN performance. Considering that the transmission capacity of optical fiber links is much higher than that of conventional WLAN with mmWave band, THz WLAN can enable seamless ultra-high-speed connections between high-speed wired networks and user mobile devices. THz beamforming provides the ability to implement bandwidth-intensive applications, such as virtual reality, high-definition holographic video conferencing, and multi-user concurrent high-speed data transmission [38]. At present, the achievable communication distance of THz is small, which drives the community to propose the idea of “information shower”, or “kiosk application”, to

### TABLE XI

| Freq (Hz) | Size                  | Process          | Beam scan | Gain   | Antenna type       | Reference |
|----------|-----------------------|------------------|-----------|--------|--------------------|-----------|
| 280 G    | 4 × 4 arrays          | 45 nm SOI CMOS   | 80°/80°   | 16 dBi | on-chip            | [247]     |
| 140 G    | 2 × 4 arrays          | 65 nm CMOS       | 40°       | -      | multi-chip         | [248]     |
| 0.53 T   | 1 × 4 arrays          | 40 nm CMOS       | 60°       | 11.7 dBi | patch              | [249]     |
| 400 G    | 1 × 8 arrays          | 45 nm SOI CMOS   | 75°       | 12 dBi | patch              | [250]     |
| 0.34 T   | 2 × 2 arrays          | 130 nm SiGe BiCMOS | 128°/53° | -      | patch              | [251]     |
| 320 G    | 1 × 4 arrays          | 130 nm SiGe BiCMOS | 24°       | 13 dBi | patch              | [252]     |
| 338 G    | 4 × 4 arrays          | 65 nm CMOS       | 45°/50°   | 18 dBi | microstrip         | [253]     |
| 317 G    | 4 × 4 arrays          | 130 nm SiGe BiCMOS | -         | 17.3 dBi | return-path gap coupler | [254]     |
| 280 G    | 2 × 2 arrays          | 65 nm CMOS       | 30°       | 12.5 dBi | dielectric resonator | [223]     |
| 300 G    | 1 × 4 arrays          | photonic         | 90°       | 10.6 dBi | bow-tie antenna    | [224]     |
| 1.05 T   | 4 × 4 arrays          | graphene         | -         | 13.9 dBi | dipole             | [264]     |
| 1.1 T    | 2 × 2 arrays          | graphene         | 60°       | 8.3 dBi | patch              | [265]     |
| 220 – 320 G | 600 elements          | metallic         | 48°       | 28.5 dBi | frequency scanning | [266]     |
| 0.8 T    | 2 × 2 arrays          | graphene         | -         | -      | photoconductive    | [272]     |
| 1.3 T    | 25448 elements        | graphene         | -         | 29.3 dBi | reflectarray       | [256]     |
| 220 – 320 G | 8 × 8 arrays          | brass sheets     | 50°/45°   | 17 dBi | frequency scanning | [273]     |
| 100 G    | 54 × 52 cells         | liquid crystals  | 55°       | 15 dBi | reflectarray       | [274]     |
| 345 G    | -                     | liquid crystals  | 20°       | 35 dBi | reflectarray       | [275]     |
| 100 G    | -                     | VO2              | 44°/44°   | -      | metasurfaces       | [237]     |
| 115 G    | 39 × 39 cells         | liquid crystals  | 20°       | 16.55 dBi | reflectarray       | [276]     |

1 In both azimuth and elevation
maximize the capabilities of T-WLAN [283]. The kiosk application recommends deploying THz access points (APs) in specific high hotspot areas (e.g., public building entrances, shopping mall halls, etc.) to provide local high-rate transmissions [284].

D. Wireless Data Center

As the demand for cloud service applications increases steadily, data centers have become an important part of modern Internet architecture. However, wired data centers result in high complexity, power consumption, maintenance costs, and large space occupied by cables. [285] proposes to introduce a wireless link in the data center to achieve higher reconfigurability and dynamic operation. THz communication technologies have the potential to provide dense and extremely fast wireless connections for data centers with high flexibility [286], [287]. Some researchers also point out that combining wireless communications with existing data center architectures can help reduce costs [288]. There have been some investigations on the THz channel modeling for wireless channel environment in data center [289]–[291].

E. Secure Transmission

The THz beamforming via ultra-massive antenna elements to create highly directional beams brings many benefits to secure transmission applications, especially in commercial scenarios and secure payment scenarios [284], [292]. To be exact, the directional beams result in fewer scattering components and space sparsity, which prevents any node located outside the beam direction from eavesdropping the information [284]. In the meanwhile, as the transmit and receive THz beams are aligned within specific angle pairs, interference from other angles can be considerably reduced.

Despite the above benefits, information security is still imperfect in the vicinity of the THz beam propagation path. Note that it is usually the case that the eavesdropper’s antenna is located in the signal propagation path to monitor the signal. In particular, [293] shows that eavesdroppers can place small (compared to the beam size) scatterers or even beam splitters on the legal receiving path to radiate legal signals to undesired positions. To tackle this challenge, [294] controls the coverage distance of the beam to improve the security of THz communication by optimizing carrier frequencies, power allocation, and rate distribution on each sub-band. In addition, it has been validated that integrating IRS to THz communication can further enhance the physical-layer security in the wiretap channel [295]–[297].

F. Networks-on-Chip Communications

Typical computer cores must constantly interact with each other to share common data and synchronize their activities. However, with the increase in the number of cores on the chip, the conventional wired topology is not enough to ensure high-speed interaction under the predicament of complex multi-core wiring [284], [298]. It proposes to use graphene-based nano-antennas to replace on-chip wiring to realize on-chip wireless communication. The chip-to-chip link can only be enabled with a smaller transceiver (sub-millimeter level). By using planar THz nano-antennas to create ultra-high-speed links, it is expected to meet the stringent requirements of on-chip scenarios with limited area and dense communications [38].
### TABLE XII
THE CHARACTERISTICS OF WIRELESS COMMUNICATION FREQUENCY BANDS.

|                      | sub-6 GHz | mmWave | THz              |
|----------------------|-----------|--------|-----------------|
| **Data rates**       | Low (below 1 Gbps) | Medium (up to 10 Gbps) | High (up to 100 Gbps) |
| **Bandwidth**        | Small (around 100 MHz) | Wide (less than 10 GHz) | Ultra-wide (up to 100 GHz) |
| **Multipath**        | Rich scattering | Few path | LoS-dominant |
| **Path loss**        | Mainly spreading loss | Spreading loss and slight molecules absorption (tens of dB/km) | Spreading loss and severe molecular absorption (100 dB/km) |
| **Angular spread**   | All around | Sparse (30 degree) | Very sparse (15 degree) |
| **Noise**            | Thermal noise | Thermal noise | Thermal noise and molecular reradiation noise |
| **Blockage effect**  | almost uninterruptible | Easily blocked | Severe blockage problem |
| **Beamwidth**        | Wide to medium | Narrow | Extremely narrow |
| **Distance**         | Big range | Medium range ≤ 200 m | Short range ≤ 50 m |
| **Number of antennas** | Limited number of elements | Small size with massive elements | Tiny size with ultra-massive elements |
| **Hardware constraint** | Normal | High | Very high |
| **Beamforming architecture** | Digital beamforming | Hybrid beamforming | Analog beamforming, TTD beamforming, IRS-assisted beamforming |

### X. OPEN CHALLENGES

In this section, we outline some open challenges of the THz beamforming in UM-MIMO systems.

#### A. Channel Modeling and Measurement

In Section II-A we consider modifying and tailoring the conventional S-V channel model to characterize the THz wave propagation. As the spatial superposition of ultra-massive single-input single-output (SISO) channels is complicated, accurate THz channel modeling is still lacking for UM-MIMO systems. Various factors need to be considered, including modeling in static and time-varying environments, channel space-time correlation under large-scale antenna arrays, near-field effects caused by the expansion of array scale, and modeling mutual coupling effects [48]. In particular, the large array size in UM-MIMO leads to inaccuracy of planar-wave channel model (PWM), while an increased channel matrix dimension leads to excessive parameters in applying spherical-wave channel model (SWM). Thus, a hybrid spherical- and planar-wave channel model (HSPM) can be adopted to balance the accuracy and overhead of channel modeling [299]. Channel measurement is an important way to verify and improve the channel model. However, channel measurement in the THz band requires high-precision equipment, diverse test scenarios, and long iteration cycles, which is associated with high costs. At present, apart from the lower THz frequencies [54], the measurements for about 1 THz is still limited [300], [301].

#### B. THz Transceiver Device

Wideband THz beamforming requires effective excitation of precise THz waves. However, the difficulty of exciting THz signals in a wide bandwidth is owing to the particularity of the THz frequency, which is too low for optical devices and exceeds the upper limit of conventional electronic oscillators. In the past ten years, some emerging solutions, e.g., SiGe, gallium nitride (GaN), indium phosphide (InP), quantum cascade laser (QCL), and graphene have been proposed to achieve efficient generation and detection of THz waves. Nevertheless, there are still many challenges in hardware devices [20]. SiGe-based devices have only limited power gain and it is difficult to operate above 500 GHz. GaN high electron mobility transistor (HEMT) faces lower breakdown voltage in some scenarios. QCL can only work in a low-temperature laboratory environment, and face the dilemma of miniaturization [53]. Graphene can be used to design compact THz transceivers due to its high conductivity and support the propagation of THz surface plasmon polariton (SPP) waves. However, graphene is far less mature than the above-mentioned technologies due to the lack of research [38].

#### C. Low-Resolution Hardware

Typically, the RF chain consists of analog-to-digital conversion/digital-to-analog conversion (ADC/DAC), demodulator, up/down-converter, low noise amplifier (LNA), mixers, automatic gain control (AGC), variable gain amplifier (VGA), and some filters. The existing signal processing algorithms developed for ideal components with infinite resolution require high-cost hardware and high energy consumption. Besides, the overall impact of non-ideal components may seriously undermine the theoretically expected performance. Thus, the investigation of low-resolution RF components is a very important topic. For example, some works have investigated the influence of nonlinear distortions, inflicted by low-resolution...
ADCs [302], [304] and quantized phase shifts [138], [305], [306], on the beamforming design.

D. Large-Scale THz Array Design

Thanks to the coherent superposition of electromagnetic waves, ultra-massive antenna elements are essential to compensate for the severe propagation loss of THz waves. Large-scale integrated phased arrays can increase the antenna gain while maintaining the advantages of miniaturization and flexible beam control. However, the practical design of a large-scale THz array is not an easy task. Challenges include the complicated signal distribution network design, the heat dissipation layout design of the dense array, and the low direct current conversion efficiency issue. As an emerging solution, graphene-based large-scale antenna arrays are expected to overcome the above issues [53]. The major challenge for graphene antennas is to characterize the interaction and coupling effects among adjacent elements [38]. Current works on graphene-based antennas still focus on theoretical analysis but lack experimental validation.

E. Wideband Beamforming

To reduce/eliminate the frequency wideband effect as discussed in Sec. VII-B, there are several reported approaches to enhance the wideband beam gain, such as using the TTD beamformer [307], redesign the codebook [308], using sub-processors for different frequency groups [212], etc. However, these technologies are of high energy consumption and require sophisticated hardware, which may not be practical in UM-MIMO systems. The overhead can be reduced by using the architecture of hybrid TTDs and phase shifters, where the joint design of the parameters on TTD and phase shifts are worthy of study in THz wideband beamforming [217], [225]. In addition, as for the beam squint effect for analog beamformer, we can use the AoSA architecture to realize wideband beamforming based on the beam split aggregation and multiplexing (BSAM) [309], which is a promising low-cost solution. For a single analog array, to enable the user to receive a wideband signal under the beam squint effect, customized wideband codebooks can be designed, where each beam serves a wide zone instead of a narrow zone as in its narrowband counterpart, so as to avoid the misalignment with the AoA due to beam squint.

F. 3D Beamforming

3D beamforming technology supported by the THz UM-MIMO system allows higher transmission gain and stronger directivity for THz communication to provide large-capacity and less-interference multiple access. This feature will help the connection management of a large number of air targets (e.g. cellular-connected drones) in the future air-ground integrated wide-area coverage network [310]. However, compared with 2D beamforming, 3D schemes bring more challenges to codebook design and beam management. On one hand, the 3D codebook directly extended from the 2D version through the Kronecker product will yield irregular coverage. It is necessary to construct a new 3D codebook to ensure the maximum coverage in both azimuth and elevation dimensions. On the other hand, extremely sharp THz beam pairs need to be continuously adjusted in high-speed mobile scenarios, which requires more efficient and fast 3D beam management. A grid-based unified beam training and beam tracking procedure can support such a fast 3D beam management [131].

G. Blockage Problem

The THz wave has limited diffraction and the LoS path is extremely easy to be blocked by obstacles [311]. When the THz waves are transmitted in the non-LoS (NLoS) channel, the data rate drops considerably due to reflection and/or scattering losses on rough surfaces [312]. Thus, one possible solution is to add APs to expand the signal coverage. However, coverage planning requires a detailed 3D model of the geographic environment to properly deploy APs, but the dense cells greatly increase the complexity of AP location planning. On the other hand, new challenges need to be fully considered before obtaining the diversity gain brought by the coordinated transmission of multiple APs, including the control signal overhead between APs and the timeliness of finding the best joint transmission strategy [313]. Another solution is adding IRSs to reconstruct the wireless environment with THz propagation [314]. Some key concerns include how to locate the IRS to enhance the signal coverage [315] and how to realize the optimal routine [316]. Besides, in high-mobility outdoor scenarios, the beam misalignment caused by Doppler expansion needs to be seriously considered [317].

XI. CONCLUSION

The THz frequency band is recently becoming a hot spot in the communication community. Table XII summarizes the features that distinguish sub-6 GHz, mmWave, and THz communications. In this paper, the model of the THz UM-MIMO systems was first established, in which the channel parameters, antenna geometry, and transceiver architecture are specified. We highlighted the basic principles of beamforming and presented the schemes of beam training. Then, we proceeded to address the on-trend THz beamforming topics, i.e., IRS-assisted joint beamforming and multi-user beamforming. For the former, we studied the model of IRS-assisted systems and provided a cooperative beam training scheme. For the latter, we provided a beamforming strategy for multi-user UM-MIMO THz systems and evaluated different linear beamforming algorithms for it. Next, we discussed the spatial-wideband and frequency-wideband effects, along with the feasible solutions in wideband beamforming. Then, the existing THz MIMO arrays were classified based on the fabrication techniques, including the electronic approach, photonic approach, and new materials approach. In the end, the emerging applications and some open challenges were elaborated. It is expected that these open problems will motivate the researchers and engineers to endeavor for more efficient solutions to future THz UM-MIMO systems.
[209] A. Tenenbaum and R. Adve, “Linear processing and sum throughput in the multilayer MIMO downlink,” IEEE Trans. Wireless Commun., vol. 8, no. 5, pp. 2652-2661, May 2009.

[210] J. Zhang et al., “Joint linear transmitter and receiver design for the downlink of multilayer MIMO systems,” IEEE Commun. Lett., vol. 9, pp. 991-993, Nov. 2005.

[211] Q. Shi et al., “An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel,” IEEE Trans. Signal Process., vol. 59, pp. 4331-4340, Sept. 2011.

[212] J. Brady and A. Sayeed, “Wideband communication with high-dimensional arrays: New results and transceiver architectures,” in 2015 IEEE Int. Conf. Commun. Workshop (ICCW), London, 2015, pp. 1042-1047.

[213] H. Hashemi et al., “Integrated true-time-delay-based ultra-wideband array processing,” IEEE Commun. Mag., vol. 46, no. 9, pp. 162-172, Sept. 2008.

[214] B. Wang et al., “Spatial- and frequency-wideband effects in millimeter-wave massive MIMO systems,” IEEE Trans. Signal Process., vol. 66, no. 13, pp. 3393-3406, July 2018.

[215] M. Cai et al., “Effect of wideband beam squint on codebook design in phased-array wireless systems,” in 2016 IEEE GLOBECOM, Washington, DC, USA, 2016, pp. 1-6.

[216] B. Wang et al., “Spatial-wideband effect in massive MIMO with application in mmWave systems,” IEEE Commun. Mag., vol. 56, no. 12, pp. 134-141, Dec. 2018.

[217] L. Dai, J. Tan, Z. Chen, and H. V. Poor, “Delay-Phase Precoding for Wideband THz Massive MIMO,” IEEE Trans. Wireless Commun., vol. 21, no. 9, pp. 7271-7286, Sept. 2022.

[218] J. Rodriguez et al., “An UWB beamformer with 4ps true time delay resolution,” in IEEE 2005 Custom Integrated Circuit Design Conf., San Jose, CA, 2005, pp. 805-808.

[219] T. Chu et al., “A 4-channel UWB beam-former in 0.13µm CMOS using a path-sharing true-time-delay architecture,” in 2007 IEEE Int. Solid-State Circuits Conf., San Francisco, CA, 2007, pp. 426-431.

[220] T. Chu and H. Hashemi, “A CMOS UWB camera with 7×7 simultaneous active pixels,” in 2008 IEEE Int. Solid-State Circuits Conf., San Francisco, CA, 2008, pp. 120-121.

[221] F. Miyama et al., “Ultrafast optical control of group delay of narrowband terahertz waves,” Sci. Rep., vol. 4, pp. 4346, 2014.

[222] Wei Jia et al., “Active control and large group delay in graphene-based terahertz metamaterials,” The J. Phys. Chem. C, vol. 123, no. 23, pp. 18560-18564.

[223] N. Buadana, S. Jameson and E. Socher, “A 280-GHz digitally controlled four-port chip-scale dielectric resonator antenna transmitter with DCD true-time delay,” IEEE Solid-State Circuits Lett., vol. 3, pp. 454-457, 2020.

[224] P. Lu et al., “Photonic assisted beam steering for millimeter-wave and THz antennas,” in 2018 IEEE Int. Conf. Antennas & Propagat., IEEE Access, vol. 6, pp. 564-580, Mar. 2018.

[225] B. Zhai and H. Hashemi, “Graphene-based plasmonic nano-antenna for terahertz band communication in nanonetworks,” IEEE J. Sel. Areas Commun., vol. 31, no. 12, pp. 685-694, Dec. 2013.

[226] B. Schrgger et al., “Discrete terahertz beam steering with an electrically controlled liquid crystal device,” J. Infrared Milli. Terahz. Waves, vol. 33, pp. 1117-1122, 2012.

[227] Correas-Serrano et al. (2017), “Graphene-based antennas for terahertz systems: A review,” [Online]. Available: https://arxiv.org/abs/1704.00371v1

[228] C. Lee et al., “Terahertz antenna arrays with silicon micromachined-microlens microantenna and corrugated horns,” in Proc. Int. Workshop Antenna Technol. (iWAT), Seoul, 2015, pp. 70-73.

[229] E. Garcia-Muñoz et al., “Photonic-based integrated sources and antenna arrays for broadband wireless links in terahertz communications,” Semicond. Sci. Technol., vol. 34, no. 5, May 2019.

[230] K. M. Luk et al., “A microfabricated low-profile wideband antenna array for terahertz communications,” Sci. Rep., vol. 7, no. 1, pp. 1268, Dec. 2017.

[231] Sengupta, T. Nagatsuma, and D. M. Mittleman, “Terahertz integrated electronic and hybrid electronic photonics systems,” Nature Electron., vol. 1, no. 12, pp. 622-635, Dec. 2018.

[232] Z. Wang et al., “A CMOS 210-GHz fundamental transceiver with OOK modulation,” IEEE J. Solid-State Circuits, vol. 49, no. 3, pp. 564-570, Mar. 2014.

[233] Z. Wang et al., “THz arrays in CMOS,” in 2020 3rd Int. Workshop on Mobile Terahertz Syst. (iWMTS), Essen, Germany, 2020, pp. 1-5.

[234] K. Sengupta and A. Hajimiri, “A 0.28 THz power-generation and beam-steering array in CMOS based on distributed active radiators,” IEEE J. Solid-State Circuits, vol. 47, no. 12, pp. 3013-3031, Dec. 2012.

[235] A. Tang et al., “A 65nm CMOS 140 GHz 27.3 dBm EIRP transmit array with membrane antenna for highly scalable multi-chip phase arrays,” in 2014 IEEE MTT-S Int. Microw. Symp. (IMS2014), Tampa, Florida, 2014, pp. 1-3.

[236] K. Guo et al., “A 0.53-THz subharmonic injection-locked phased array with 63-µW radiated power in 40-nm CMOS,” IEEE J. Solid-State Circuits, vol. 54, no. 2, pp. 380-391, Feb. 2019.

[237] Y. Yang et al., “An eight-element 370-410-GHz phased-array transmitter in 45-nm CMOS SOI with peak EIRP of 8-8.5 dBm,” IEEE Trans. Microw. Theory Techn., vol. 64, no. 12, pp. 4241-4249, Dec. 2016.

[238] H. Jilli and O. Momeni, “A 0.34-THz wideband-angle 2-D steering phased array in 0.13-μm SiGe BiCMOS,” IEEE J. Solid-State Circuits, vol. 54, no. 9, pp. 2449-2461, Sept. 2019.

[239] X. Deng et al., “A 320-GHz 1×4 fully integrated phased array transmitter using 0.13-μm SiGe BiCMOS technology,” IEEE J. Solid-State Circuits, vol. 50, no. 2, pp. 597-609, Feb. 2015.

[240] R. Han et al., “A SiGe terahertz heterodyne imaging transmitter with 3.3 mW radiated power and fully-integrated phase-locked loop,” IEEE J. Solid-State Circuits, vol. 50, no. 12, pp. 2935-2947, Dec. 2015.

[241] M. Che et al., “Optoelectronic THz-wave beam steering by arrayed photomixers with integrated antennas,” IEEE Photon. Technol. Lett., vol. 32, no. 16, pp. 1797-1800, Aug. 2020.

[242] E. Carrusco and I. Perruisseau-Carrier, “Reflectarray antenna at terahertz using graphene,” IEEE Antennas Wireless Propag. Lett., vol. 12, pp. 253-256, 2013.

[243] Y. Luo et al., “Graphene-based multi-beam reconfigurable THz antennas,” IEEE Access, vol. 7, pp. 30802-30808, 2019.
in IEEE Radio Freq. Integr. Circuits Symp. (RFIC), Philadelphia, PA, USA, June 2018, pp. 272-275.

[308] M. Cai, J. N. Laneman, and B. Hochwald, “Beamforming codebook compensation for beam squint with channel capacity constraint,” in IEEE Int. Symp. Inf. Theory, June 2017, pp. 76-80.

[309] B. Ning, L. Li, W. Chen, and Z. Chen, “Wideband Terahertz Communications with AoSA: Beam Split Aggregation and Multiplexing,” in Global Commun. Conf., Rio de Janeiro, Brazil, 2022, pp. 1-6.

[310] L. Liu et al., “Multi-beam UAV communication in cellular uplink: Cooperative interference cancellation and sum-rate maximization,” IEEE Trans. Wireless Commun., vol. 18, no. 10, pp. 4679-4691, Oct. 2019.

[311] C. Jansen et al., “The impact of reflections from stratified building materials on the wave propagation in future indoor terahertz communication systems,” IEEE Trans. Antennas Propag., vol. 56, no. 5, pp. 1413-1419, May 2008.

[312] C. Jansen et al., “Diffuse scattering from rough surfaces in THz communication channels,” IEEE Trans. THz Sci. Technol., vol. 1, no. 2, pp. 462-472, Nov. 2011.

[313] L. You et al., “Network massive MIMO transmission over millimeter-wave and terahertz bands: Mobility enhancement and blockage mitigation,” IEEE J. Sel. Areas Commun., vol. 38, no. 12, pp. 2946-2960, Dec. 2020.

[314] B. Ning, P. Wang, L. Li, Z. Chen and J. Fang, “Multi-IRS-Aided Multi-User MIMO in mmWave/THz Communications: A Space-Orthogonal Scheme,” IEEE Trans. Commun., vol. 70, no. 12, pp. 8138-8152, Dec. 2022.

[315] S. Zeng et al., “Reconfigurable intelligent surface (RIS) assisted wireless coverage extension: RIS orientation and location optimization,” IEEE Commun. Lett., vol. 25, no. 1, pp. 269-273, Jan. 2021.

[316] W. Mei and R. Zhang, “Cooperative beam routing for multi-IRS aided communication,” IEEE Wireless Commun. Lett., vol. 10, no. 2, pp. 426-430, Feb. 2021.

[317] L. You et al., “BDMA for millimeter-wave/terahertz massive MIMO transmission with per-beam synchronization,” IEEE J. Sel. Areas Commun., vol. 35, no. 7, pp. 1550-1563, July 2017.

Boyu Ning received the B.E. degree in Communication Engineering, along with the Certification of the Talent Program in Yingcai Honors College, from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2018. He is currently pursuing the Ph.D. degree with the National Key Laboratory of Science and Technology on Communications, UESTC. He was a recipient of the Honor of Outstanding Students of UESTC. His research interests include Terahertz communication, intelligent reflecting surface, massive MIMO, physical-layer security, and convex optimization.

Zhongbao Tian received the B.E. degree in communication engineering from the University of Electronic Science and Technology of China (UESTC) in 2019. He is currently working toward the M.E. degree with the National Key Laboratory of Science and Technology on Communications, UESTC. His research and study interests include Terahertz communication and 3D beam forming.

Weidong Mei received the B.Eng. degree in communication engineering and the M.Eng. degree in communication and information systems from the University of Electronic Science and Technology of China, Chengdu, China, in 2014 and 2017, respectively, and the Ph.D. degree from the NUS Graduate School, National University of Singapore, in 2021 under the Integrative Sciences and Engineering Programme (ISEP) Scholarship. From 2021 to 2022, he was a Research Fellow with the Department of Electrical and Computer Engineering, National University of Singapore. He is currently a Professor with the University of Electronic Science and Technology of China. His research interests include intelligent reflecting surface, wireless drone communications, physical-layer security, and convex optimization techniques. He was a recipient of the Outstanding Master’s Thesis Award from the Chinese Institute of Electronics in 2017, and the Best Paper Award from the IEEE International Conference on Communications in 2021. He was honored as the Exemplary Reviewer of the IEEE TRANSACTIONS ON COMMUNICATIONS from 2019 to 2020, the IEEE WIRELESS COMMUNICATIONS LETTERS in 2019 and 2021, and the IEEE COMMUNICATIONS LETTERS from 2021 to 2022.

Zhi Chen received B. Eng, M. Eng., and Ph.D. degree in Electrical Engineering from University of Electronic Science and Technology of China (UESTC), in 1997, 2000, 2006, respectively. On April 2006, he joined the National Key Lab of Science and Technology on Communications (NCL), UESTC, and worked as a professor in this lab from August 2013. He was a visiting scholar at University of California, Riverside during 2010-2011. He is also the deputy director of Key Laboratory of Terahertz Technology, Ministry of Education. His current research interests include Terahertz communication, 5G/6G mobile communications.

Chong Han is with Shanghai Jiao Tong University, Shanghai, China, where he is currently a John WU & Jane Sun Endowed Associate Professor, and the Director of the Terahertz Wireless Communications (TWC) Laboratory. He obtained the Ph.D. degree in Electrical and Computer Engineering from Georgia Institute of Technology, Atlanta, GA, USA, in 2016. He received 2019–2021 Distinguished TPC Member Award, IEEE International Conference on Computer Communications (INFOCOM) and 2018 Elsevier Nano Communication Network Journal Young Investigator Award, 2018 Shanghai Chenguang Funding Award, and 2017 Shanghai Yangfan Funding Award. He is a (guest) editor of IEEE JSAC, JSTSP, Open Journal of Vehicular Technology, etc. He has published 6 book chapters, 50+ journal articles, and 60+ conference papers, most of which, if not all, are related to THz communications. He is a TPC Co-Chair or General Co-Chair for the 1st–6th International Workshop on Terahertz Communications, in conjunction with IEEE ICC/Globecom. Furthermore, he will be the symposium chair of SAC THz Communications with Globecom’2023. He is serving as a Vice Chair of IEEE ComSoc RCC Special Interest Group (SIG) on THz Communications.
Shaoqian Li (Fellow, IEEE) received the B.E. degree in communication technology from Northwest Institute of Telecommunication Engineering (currently Xidian University), Xian, 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. He joined UESTC as an Academic Member in 1984, where he became a Professor of information and communication systems in 1997, and a Ph.D. Supervisor in 2000. He is currently the Director of the National Key Laboratory of Communications, UESTC. He has authored hundreds of journal or conference papers, and published several books. His research topics cover a broad range, including multiple-antenna signal processing technologies for mobile communications, cognitive radios, coding and modulation for next generation mobile broadband communications systems, wireless and mobile communications, anti-jamming technologies, and signal processing for communications. He has been a member of the Communication Expert Group of the National 863 Plan since 1998 and a member of The Future Project since 2005. He was the TPC Co-Chair of the IEEE International Conference on Communications, Circuits, and Systems in 2005, 2006, and 2008. He is currently a member of the Board of Communications and Information Systems of Academic Degrees Committee, State Council, China. He is also a member of the Editorial Board of the Chinese Science Bulletin and the Chinese Journal of Radio Science.

Jinhong Yuan (M’02–SM’11–F’16) received the B.E. and Ph.D. degrees in electronics engineering from the Beijing Institute of Technology, Beijing, China, in 1991 and 1997, respectively. From 1997 to 1999, he was a Research Fellow with the School of Electrical Engineering, University of Sydney, Sydney, Australia. In 2000, he joined the School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, Australia, where he is currently a Professor and Head of Telecommunication Group with the School. He has published two books, five book chapters, over 300 papers in telecommunications journals and conference proceedings, and 50 industrial reports. He is a co-inventor of one patent on MIMO systems and four patents on low-density-parity-check codes. He has co-authored four Best Paper Awards and one Best Poster Award, including the Best Paper Award from the IEEE International Conference on Communications, Kansas City, USA, in 2018, the Best Paper Award from IEEE Wireless Communications and Networking Conference, Cancun, Mexico, in 2011, and the Best Paper Award from the IEEE International Symposium on Wireless Communications Systems, Trondheim, Norway, in 2007. He is an IEEE Fellow and currently serving as an Associate Editor for the IEEE Transactions on Wireless Communications and IEEE Transactions on Communications. He served as the IEEE NSW Chapter Chair of Joint Communications/Signal Processings/Ocean Engineering Chapter during 2011-2014 and served as an Associate Editor for the IEEE Transactions on Communications during 2012-2017. His current research interests include error control coding and information theory, communication theory, and wireless communications.

Rui Zhang (S’00–M’07–SM’15–F’17) received the B.Eng. (first-class Hons.) and M.Eng. degrees from the National University of Singapore, Singapore, and the Ph.D. degree from the Stanford University, Stanford, CA, USA, all in electrical engineering.

From 2007 to 2009, he worked as a researcher at the Institute for Infocomm Research, ASTAR, Singapore. In 2010, he joined the Department of Electrical and Computer Engineering of National University of Singapore, where he was appointed as a Provost’s Chair Professor in 2020. Since 2022, he has joined the School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, as a Principal’s Dilegence Chair Professor. He has published over 300 journal papers and over 200 conference papers. He has been listed as a Highly Cited Researcher by Thomson Reuters/Clarivate Analytics since 2015. His current research interests include UAV/satellite communications, wireless power transfer, intelligent reflecting surface, reconfigurable MIMO, radio mapping and optimization methods.

He was the recipient of the 6th IEEE Communications Society Asia-Pacific Region Best Young Researcher Award in 2011, the Young Researcher Award of National University of Singapore in 2015, the Wireless Communications Technical Committee Recognition Award in 2020, and the IEEE Signal Processing and Computing for Communications (SPCC) Technical Recognition Award in 2020. He received 14 IEEE Best Paper Awards, including the IEEE Marconi Prize Paper Award in Wireless Communications in 2015 and 2020, the IEEE Signal Processing Society Best Paper Award in 2016, the IEEE Communications Society Heinrich Hertz Prize Paper Award in 2017, 2020 and 2022, the IEEE Communications Society Stephen O. Rice Prize in 2021, etc. He served for over 30 international conferences as the TPC co-chair or an organizing committee member. He was an elected member of the IEEE Signal Processing Society SPCOM Technical Committee from 2012 to 2017 and SAM Technical Committee from 2013 to 2015, and served as the Vice Chair of the IEEE Communications Society Asia-Pacific Board Technical Affairs Committee from 2014 to 2015. He was a Distinguished Lecturer of IEEE Signal Processing Society and IEEE Communications Society from 2019 to 2020. He served as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2012 to 2016, the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS: Green Communications and Networking Series from 2015 to 2016, the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2013 to 2017, the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING from 2016 to 2020, and the IEEE TRANSACTIONS ON COMMUNICATIONS from 2017 to 2022. He served as a member of the Steering Committee of the IEEE Wireless Communications Letters from 2018 to 2021. He is a Fellow of the Academy of Engineering Singapore.