Prospective Beamforming Technologies for Ultra-Massive MIMO in Terahertz Communications: A Tutorial

Boyu Ning, Student Member, IEEE, Zhongbao Tian, Student 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 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 tutorial on the 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 introduce the schemes of beam training and beamspace MIMO for THz communications. Moreover, the spatial-wideband effect and frequency-wideband effect in the THz beamforming are introduced. The intelligent-reflecting-surface (IRS)-assisted beamforming and multi-user beamforming in THz UM-MIMO systems are discussed, respectively. 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-input multiple-output (MIMO) 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 (Het-Nets) [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.

To compensate for the propagation loss, the ultra-massive MIMO (UM-MIMO) systems, which can generate high-gain directional beams via beamforming technologies, are considered a pragmatic solution to be integrated into THz communications. In this paper, 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 shed light on the current research status and the progress of THz UM-MIMO. Finally, the emerging applications and open challenges are elaborated. This paper is not meant to be a comprehensive survey of an existing subject, but rather serves as a tutorial 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 1 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 [21]. In 2004, M. J. Fitch and R. Osiander further discussed the sources, detectors, and modulators for practical THz systems [22]. In 2007, several magazines came up with reports of THz technology progress status and applications of THz systems [23–25]. 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 [26]. After that, K. C. Huang and Z. Wang provided a tutorial on constructing robust, low-cost THz wireless systems, in 2011 [27]. 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 [28–30]. 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 [31], [32]. In 2014, T. Kurner and S. Priebe reported the current research projects, spectrum regulations, and ongoing standardization activities in THz communication systems [33]. 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 [37]. In 2015, A. Hirata and M. Yaita gave a brief overview of the THz technologies and standardization of wireless communications [38]. 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 [39]. 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 [40], [41]. J. F. Federici et al. gave a survey to illustrate the impact of weather on THz wireless links [42]. In 2017, S. Muntaz et al. discussed the opportunities and challenges in THz communications for vehicular networks [43]. In 2018, V. Petrov et al. provided a tutorial on propagation modeling, antenna, and testbed designs [44]. A. Boulogeorgos et al. reported the basic system architecture for THz wireless links with bandwidths into optical networks [45]. C. Han and Y. Chen reported three methods, e.g., deterministic, statistical, and hybrid methods, to model THz propagation channels [46]. I. F. Akyildiz et al. focused on the solution to the THz distance limitation [47]. N. Khalid et al. provided a tutorial on performing THz modulation schemes [48]. D. Headland [49] 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 for THz communications [20], [50–55]. In 2020, H. Saried-deen et al. reported the current THz technologies in wireless communications, imaging, sensing, and localization [56]. L. Zhang et al. reported the key technologies of optoelectronic THz communications in the physical layer [57]. M. A. Jamshed et al. conducted a survey on the antenna selection design for THz applications [58]. S. Ghafoor et al. gave an overview of THz media access control (MAC) protocols with classifications, band features, design issues, and future challenges [59]. C. X. Wang et al. reported the channel measurements, characteristics, and models for THz frequency band [60]. A. Faisal et al. reported the advantages of UM-MIMO systems in THz communications and discussed the challenges and shortcomings [61]. In 2021, F. Lemic et al. conducted a comprehensive survey.
| 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. Piesiewicz, et al.   | 2007 | Short-Range Ultra-Broadband Terahertz Communications: Concepts and Perspectives | Magazine   | This article reports the concept of ultra-broadband THz communications 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 paper 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 communica tions | 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 commun ications: 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. Akvildiz 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. Akvildiz 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. Mumtaz 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. Akvildiz 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 band 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. Tekbyak et al.       | 2019 | Terahertz band communication systems: Challenges, novelties 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. Huq et al.     | 2019 | Terahertz-Enabled Wireless System 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. Elavan et al.  | 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. Sarieddeen 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.   | 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.  | 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.     | 2020 | THz precoding for 6G: Applications, challenges, solutions, and opportunitiies | 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.   | 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.     | 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. Sarieddeen 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. |
| C. Chaccour et al. | 2021 | 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. |
| S. Tripathi et al. | 2021 | Millimeter-wave and Terahertz Spectrum for 6G Wireless                  | Survey     | This paper gives an overview of mmWave and THz communications from wireless propagation characteristics, channel models, implementations, potential applications, and standardization activities. |
| H. J. Song        | 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.      | 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.    | 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.    | 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. |
on THz nano-communication and network from a top-down perspective [63]. C. Han et al. investigated the architectures and challenges of hybrid beamforming in THz communications [64]. H. Sarieddeen [65], C. Chaccour [66], and S. Tripathi [67] gave overviews of signal processing techniques, propagation characteristics, and features 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].

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 [72]. In the above-mentioned works, the authors in [61] provided a brief overview of THz UM-MIMO systems and highlighted a few challenges. The authors in [46] 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 Tutorial

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 via UM-MIMO and discuss the effects of near-field and far-field propagation. Furthermore, we introduce two beamforming schemes, i.e., beam training and beamspace MIMO, which are considered promising and feasible solutions for THz communications.
- 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 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 the inter-user interference and evaluate their performance in THz UM-MIMO systems.
- 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 terms of channel modeling and measurement, THz transceiver device, low-resolution hardware, large-scale THz array design, 3D beamforming, and mobility versus blockage trade-off, to light up new horizons and stimulate enthusiasm for future research.
Table II: Three Evaluation Models for the Attenuation of Electromagnetic Wave between 0.1 and 1 THz.

| Model                | Description                                                                 |
|----------------------|-----------------------------------------------------------------------------|
| MPM model [83]       | 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 [84]        | AM is a tool for radiative transfer computations at the microwave to submillimeter 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 [85]     | ITU-R E676-11 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. |

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 wideband beamforming and the IRS-assisted joint beamforming in THz UM-MIMO systems. Section VI discusses the multi-user beamforming schemes in THz UM-MIMO systems. Section VII provides an overview of the existing THz MIMO arrays. Section VIII identifies the emerging applications with THz beamforming technologies. Open challenges and future research directions are discussed in Section IX. Finally, we conclude the paper in Section X.

Notation: We use small normal face for scalars, small bold face for vectors, and capital bold face for matrices. The superscript \( \{ \cdot \}^T \), \( \{ \cdot \}^\dagger \), 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.

II. THz UM-MIMO Systems

In this section, we start with describing the model and specifying the parameters of THz UM-MIMO systems. Then, some important system characteristics are elaborated, i.e., antenna geometry 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 signals \( y \in \mathbb{C}^{N_r} \) at the \( N_r \) antenna elements can be expressed as

\[
y = Hx + n, \tag{1}
\]

where \( x \in \mathbb{C}^{N_t} \) is the transmitted signals at the \( N_t \) antenna elements after the beamforming (also referred to as precoding). \( H \) is the THz UM-MIMO channel and \( n \) is the noise vector. The THz wireless channel modeling can be realized by deterministic [73], [74], statistical [75], and hybrid approaches [76], [77]. In this paper, we provide a basic channel model as follows.

With the employment of ultra-massive antenna elements, the THz channel generally shows sparsity and strong directivity, which is composed of one or more propagation paths with different angles of departure/arrival (AoD/AoA) [78]–[80]. The measurement and modeling works for THz channels have been reported in [81]–[85]. Based on the Saleh-Valenzuela (S-V) channel model [86] and the measurement results, the channel matrix at time \( t \) can be explicitly written as

\[
H(t) = \sum_{i=1}^{N_c} \sum_{j=1}^{N_{ray}} \delta(t - \tau_{ij}) \sqrt{\alpha_{ij} G_{t,ij} G_{r,ij} a_r a_t^H}, \tag{2}
\]

where \( N_c \) 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 \( T_1 = 0 \)) and \( T_{ij} \) (with \( T_1 = 0 \)) represent the arrive time of the \( i \)-th cluster and that of the \( j \)-th ray in it, respectively. \( G_{t,ij} \) and \( G_{r,ij} \) (resp. \( a_r \) and \( a_t \)) 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.

1) Path Gain: For mmWave channel models, a zero mean, unit variance circularly symmetric complex Gaussian distributed path gain is widely assumed in works about signal processing [87]–[91]. However, the path gain of THz channels suffers from severe free spreading loss due to the extremely high frequency. According to the Friis’ formulation [92], the free spreading loss is given by

\[
L_{\text{spread}}(f, d) = \left( \frac{4\pi f d}{c} \right)^2, \tag{3}
\]

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

\[
\alpha_{ij}(f, d_{ij}) = \frac{1}{L_{\text{spread}}(f, d_{ij}) L_{\text{abs}}(f, d_{ij})} e^{-\frac{\tau_{ij}}{\tau_{ij}}} e^{-\frac{\tau_{ij}}{\tau_{ij}}}, \tag{4}
\]

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...
By using the molecular absorption coefficient shown in Fig. 3, future works.

For ease of access, we specify promising to be used for wireless communication in the future.

It is observed from the red line in Fig. 3 that there are some frequencies above which the total absorption coefficient is relatively large at the center points of these frequencies, albeit with the difference of the order of magnitude. Except for these jump points, the path gain calculated by our proposed method is accurate and reliable.

2) Antenna Gain: The transmit/receive antenna gain (composed of two parts: element gain and array gain) is closely related to the element number, element distribution, and beam direction. Let \( G_e(\varphi, \theta) \) represent the element gain for a propagation path with azimuth angle \( \varphi \) and elevation angle \( \theta \) to the array in the \( xy \)-plane. \( G_e(\varphi, \theta) = \zeta F(\varphi, \theta)D \) describes how much power is transmitted/received compared to an isotropic antenna, in which

\[
D = \frac{4\pi}{\int_{\varphi=0}^{2\pi} \int_{\theta=0}^{\pi} F(\varphi, \theta) \sin \theta d\theta d\varphi}
\]

is the maximum directivity; \( \zeta \) is the antenna efficiency; \( F(\varphi, \theta) \) is the normalized radiation pattern. In practice, the element gain is focused more in the direction perpendicular to the element. An exemplary radiation pattern shown in \[23\] is given by \[7\] and also illustrated in Fig. 5.

![Fig. 3: \( k(f) \) at frequencies from 0.1 THz to 10 THz (temperature \( T=296 \) K and pressure \( P=1 \) atm in sunny day).](image)

**TABLE III**

| 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 |

from water vapor \[82\]. \( \Gamma_e \) and \( \Gamma_r \) are the exponential attenuation factors of the arrive cluster and ray, respectively, which are frequency and wall-material dependent in general \[86\].

The molecular absorption at the frequency below 1 THz can be evaluated by the atmospheric mmWave propagation model (MPM) \[83\], atmospheric model (AM) \[84\], and ITU-R P.676-10 model \[85\] for various environments. A brief introduction of these models is relegated to Table III. In general, the molecular absorption can be expressed as

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

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 \[93\]. 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 prospective THz communications are generally considered below 1 THz since the spreading loss is unaffordable in the band above it. It is observed from the red line in Fig. 3 that there are some frequency bands with smaller absorption, which hopefully are promising to be used for wireless communication in the future.

For ease of access, we specify \( k(f) \) at the center points of these candidates in Table III. These results are expected to provide rational and equitable validation, evaluation, and simulation in future works.

Let us consider the total path loss of the LoS channel in \[2\]. By using the molecular absorption coefficient shown in Fig. 3 based on \[5\], we can obtain the exact path gain of \( \alpha_{11}(f, d_{11}) \) in \[4\]. By this means, we select \( k(f) \) within 0.1-1 THz and compute the corresponding path loss based on \[4\] and \[5\]. 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.

![Fig. 4: The path loss computed by two methods.](image)
Based on (9), it is worth mentioning that the element radiation pattern (i.e., element gain) limits the gain of the total array (i.e., antenna gain). Since we cannot control the element gain, it is power spectral density (PSD) can be expressed as \[98\] frequency but independent of the transmitted signal power. Its parts, i.e., the electronic thermal noise and the re-radiation noise from each element, i.e., 

\[
G_{r,ij}[\text{dB}] = \frac{10 \log_{10}(N_t)}{10},
\]

\[
G_{r,ij}[\text{dB}] = \frac{10 \log_{10}(N_r)}{10},
\]

return that some works \[99\]–\[101\] assumed that \(\eta_{ij} = 1\), \(\forall ij\). In fact, the molecular absorption happens everywhere along with the propagation, and the spread of the re-radiated power is omnidirectional, which will not be all captured by the receiver. Regarding this, \(\eta_{ij}\) should be rather small and need to be properly modeled.

**B. Antenna Geometry**

The array response vectors \(\mathbf{a}_{r,ij}\) in (2) are functions related to the antenna geometry, which represent the phase difference on each elements \[102\]. Generally speaking, there are four typical array structures: ULA, uniform rectangular planar array (URPA), uniform hexagonal planar array (UHPA), and uniform circular planar array (UCPA).

Considering an \(N_a\)-element ULA, as depicted in Fig. 7 (a), the array response vector can be written as

\[
\mathbf{a}(\varphi)_{ULA} = \frac{1}{\sqrt{N_a}} \left[1, e^{j k d_x \cos(\varphi)}, \ldots, e^{j k (N_x - 1)d_x \cos(\varphi)}\right]^T,
\]

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

Considering an URPA with \(N_y\) times \(N_z\) elements lying on the \(yz\)-plane, as shown in Fig. 4 (b), the array response vector is given by

\[
\mathbf{a}(\varphi, \theta)_{URPA} = \frac{1}{\sqrt{N_a}} \left[1, e^{j k m d_y \sin(\theta) \sin(\varphi) + n d_x \cos(\theta)}\right.\]

\[
\vdots, e^{j k ((N_y - 1)d_y \sin(\theta) \sin(\varphi) + (N_z - 1)d_z \cos(\theta))}\right]^T,
\]

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 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.

The key factors that determine the downlink budget are base transmit power, transmitter and receiver antenna size, and path loss.

As a typical array element arrangement, the URPA structure has been widely used in THz communication scenarios. Table IV shows the link budget details for three different THz application cases using URPA.

Considering a UHPA consisting of $V$ hexagon rings on the $xy$-plane, as shown in Fig. 7(c). 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

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

where $\{f_v\}_{v=-V}^{V}$ denote the ULA vectors at different rows, $N_a = 1 + \sum_{v=1}^{V} 6v$, is the total number of elements. For different parities of the subscript $v$, $f_v$ can be expressed as

$$f_v = e^{-jvd_x \sin(\theta) \cos(\varphi)} \left[ e^{-j(V-\frac{1}{2})kd_x \sin(\theta) \cos(\varphi)}, \ldots, e^{-jkd_x \sin(\theta) \cos(\varphi)}, 1, e^{jkd_x \sin(\theta) \cos(\varphi)}, \ldots, e^{j(V-\frac{1}{2})kd_x \sin(\theta) \cos(\varphi)} \right],$$

$$f_v = e^{-jvd_y \sin(\theta) \cos(\varphi)} \left[ e^{-j(V-\frac{1}{2})kd_y \sin(\theta) \cos(\varphi)}, \ldots, e^{-jkd_y \sin(\theta) \cos(\varphi)}, 1, e^{jkd_y \sin(\theta) \cos(\varphi)}, \ldots, e^{j(V-\frac{1}{2})kd_y \sin(\theta) \cos(\varphi)} \right].$$

Considering a UCPA consisting of $C$ circles on the $xy$-plane, each element uniformly distributed over the circle as shown in Fig. 7(d), the array response vector is given by

$$a(\varphi, \theta)_{UCPA} = \frac{1}{\sqrt{N_a}} \left[ 1, \ldots, e^{jkr_c \sin(\theta) \cos(\varphi - \varphi_{t,c})}, \ldots, e^{jkr_c \sin(\theta) \cos(\varphi - \varphi_{t,C})} \right]^T,$$

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, $\varphi_{t,c}$ is the angle the $n$-th element in the $c$-th circle to the $x$-axis.

The distribution of the antenna elements, i.e., the array topology, is of interest in beamforming due to the following facts: (i) Different distribution leads to different antenna hardware packaging capacities. (ii) Different element distribution shows different spatial-correlation effects in UM-MIMO systems, (iii) The planar arrays, e.g., URPA, UHPA, and

| Cases                  | Kiosk download (dBm) | Data center (dBm) | Fronthaul/backhaul (dBm) |
|------------------------|----------------------|-------------------|--------------------------|
| Frequency (THz)        | 0.14                 | 0.22              | 0.30                     |
| Bandwidth (GHz)        | 5.00                 | 20.00             | 50.00                    |
| Distance (m)           | 10.00                | 50.00             | 200.00                   |
| Propagation loss (dB)  | –95.36               | –113.27           | –128.00                  |
| Molecules absorption (dB) | –0.0026               | –0.024            | –0.44                    |
| Atmospheric attenuation (dB) | –174                 | –174              | –174                     |
| Tx antenna size        | 16 × 16              | 32 × 32           | 32 × 32                  |
| Tx power (dBm)         | 3.00                 | 10.00             | 25.00                    |
| Tx antenna gain (dB)   | 24.00                | 30.00             | 30.00                    |
| Rx antenna size        | 4 × 4                | 32 × 32           | 32 × 32                  |
| Rx antenna gain (dB)   | 12.00                | 30.00             | 30.00                    |
| Noise figure (dB)      | 8.00                 | 8.00              | 8.00                     |
| Other losses (dB)      | 10.00                | 10.00             | 10.00                    |
| SNR (dB)               | 2.52                 | 9.70              | 3.57                     |
| Data rate (Gbps)       | 7.39                 | 67.36             | 85.59                    |
UCPA, are possible to enable 3D beamforming.

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 \[105\]. Since the phase shifters can not change the magnitude of the symbol, the analog beamforming vector is subject to constant modulus constraint, which limits the flexibility of control and impairs the beamforming performance and capacity improvement. In contrast, as a high-cost architecture, digital beamforming or precoding can realize any linear transformation of multiple signal streams from the digital baseband to the antenna elements, which provides more degree of freedom (DoF) \[106\]. Despite the ease of beam control, it is unaffordable to be applied in UM-MIMO systems due to the high power consumption and high system cost \[107\].

As a cost-performance trade-off, hybrid beamforming (HB) architecture has emerged as an attractive solution for UM-MIMO systems \[108\], \[109\]. 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 \[110\], \[111\]. There are mainly three types of HB architectures that have been reported extensively: the fully-connected architecture, the partially-connected or sub-connected architecture, and the dynamically-connected architecture.

1) Fully-Connected Architecture: In the fully-connected architecture equipped with $N_{RF}$ RF chains and $N_t$ antenna elements, as shown in Fig. 9(a), 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

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

(18)

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 length of baseband transmitted symbols. The elements of each column in matrix $\mathbf{F}_{AB}$ are phase shifters connected by each RF chain, while the elements in each row are phase shifters connected to each 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.) \[112\].

2) Partially-Connected Architectures: As shown in Fig. 9 (b), in partially-connected architecture, each RF chain is connected to a specific subset of antenna elements \[113\], i.e., a subarray. With this architecture, $\mathbf{F}_{AB}$ in (18) has the form of a block diagonal matrix as $\mathbf{F}_{AB} = \text{diag}[\mathbf{f}_1, \ldots, \mathbf{f}_{N_{RF}}]$, where $\{\mathbf{f}_i\}_{i=1}^{N_{RF}} = \frac{N_s}{N_{RF}} \times 1$ phase shifter vectors connected to the RF chains. $\mathbf{F}_{AB}$ can also be expressed as

$$\mathbf{F}_{AB} = \Phi_{AB}\mathbf{M},$$

(19)

where $\mathbf{M}$ can be regarded as a switch matrix of dimension $N_t \times N_{RF}$, $\mathbf{m}_i$ 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 shift 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 \[112\], \[114\], which is more appealing in THz communications. For example, the authors in \[115\], \[116\] implemented the partially-connected architecture for the beamforming in THz communications. In addition, the authors in \[117\] proposed a novel partially-connected HB architecture.

---

**Fig. 8.** Illustration of the hybrid beamforming that combines the digital and analog beamforming.
with two digital beamformers, wherein the additional one is developed to compensate for the performance loss caused by the analog constraints as well as the difference of channel matrices among subcarriers.

3) Dynamically-Connected Architecture: One disadvantage of the partially-connected architectures is the fixed circuit connection, which prevents adaptive and dynamic control \[118\]. To improve the flexibility and DoF in designing the analog domain, a dynamically-connected HB architecture was reported in \[118\]–\[121\]. Dynamically-connected architecture is given in Fig. 9(c), where a switching network \(W_S\) is added between the RF chains and the antenna elements \[118\], and the transmitted signal can now be rewritten as

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

where \(\Phi_{AB}\) follows the definition in \[19\], \(W_S = \{w_{i,j}\}, i = 1, \ldots, N_t, j = 1, \ldots, N_{RF}\), 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 a 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 the processing freedom through the flexible switch network, which can be regarded as the transition architecture from full connection to partial connection.

Generally, the dynamically-connected architecture refers to the above-mentioned structure, in which one antenna is equipped with one phase shifter. However, the authors in \[122\] proposed a dynamic array-of-subarrays architectures (DAoSA) with one antenna connected to \(N_{RF}\) phase shifters, claiming to

| Beamforming architectures | Advantages | Limitations | Summarize |
|---------------------------|------------|-------------|----------|
| Digital                   | Optimal performance, flexible space division multiplexing, high DoF. | Lots of RF chains, expensive hardware costs. | The best performance, the worst feasibility, not suitable for the THz band. |
| Analog                    | Simpler circuits, consume less power, low cost, easy to implement. | Lots of phase shifters, limited by the quantized phase shifter, only provide single stream transmission, poor performance. | Low cost, limited performance, feasible in the THz band. |
| HB                        | Lower power consumption than the all-digital architecture, better multi-stream transmission performance than analog architecture. | Hardware complexity and the power consumption are unacceptably high, performance is limited by the number of RF chains. | A compromise between hardware complexity and performance, balance between power consumption and data rate. |
| FC                        | Easy to expand to distributed M-MIMO. | Performance loss, sacrificing the data rate performance. | |
| PC                        | Fewer phase shifters than FC, higher algorithm design flexibility than PC. | Too many switches, high algorithm design complexity. | |
| TTD                       | Avoid beam squint effects. | Consumes more power than PSs. | Suitable for broadband beamforming. |

Fig. 9. Three types of hybrid beamforming architecture: (a) fully-connected (FC); (b) partially-connected (PC); (c) dynamically-connected (DC).
save power compared to the fully-connected architecture. From the perspective of hardware structure, DAoSA is equivalent to adding some switches to the fully-connected HB architecture but its performance does not exceed the fully connected HB. We summarized the advantages and limitations of various beamforming architectures implemented in the THz communications in Table VI.

### 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. In low-frequency and small-scale MIMO schemes, beamforming is optimized by leveraging the channel information. However, this way is no longer feasible for THz UM-MIMO systems due to i) its severe path loss and ii) unaffordable processing time for the ultra-massive array. In this context, we introduce two concepts, i.e., beam training and beamspace MIMO, which are envisioned as viable beamforming schemes in THz UM-MIMO communications.

#### A. Nomenclature

Considering a point-to-point system in (1), the received signal after spatial post-processing can be expressed as $$\hat{y} = W^H F s + W^H n$$, in which $F \in \mathbb{C}^{N_t \times N_r}$ is the beamformer and $W \in \mathbb{C}^{N_r \times N_s}$ is the combiner. The processing from the data stream $s$ to the transmit signals $x = Fs$ is called beamforming. On the contrary, the processing from the received signals $y$ to the data stream $\hat{y} = W^H y$ 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 this case of combining is named decoding. The hardware 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. In other words, digital beamforming is equivalent to precoding, **beam steering** is referred to as a special case of analog beamforming with $N_s = 1$.

Commonly, in beam steering schemes, there is a predefined codebook for transmitter/receiver, where the system sweeps the codewords and selects one beam pair with the maximum beam gain for sensing or data transmission. For example, the set of array response vectors is a classic codebook to realize directional narrow beams, which is widely used in radar systems. We summarized the features of precoding/decoding, beamforming/combining, and beam steering in Table VI.

#### 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 [123]. 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 canceling to suppress radiation in undesired directions. If phased arrays are employed at both transmitter and receiver, the resulting signal can be written as

$$y = w^H H f s + w^H n,$$

where $f \in \mathbb{C}^{N_r \times 1}$ is the beamformer with $N_r$ phase shifters, $w \in \mathbb{C}^{N_t \times 1}$ is the combiner with $N_t$ 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., $|f(i)| = \sqrt{\frac{P}{N_r}}$ 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.

We consider a LoS communication by assuming only one direct path in the THz channel, i.e., $\text{rank}(H) = 1$. Based on the channel model given in (2), it is easy to prove that any signal at the receive antenna elements 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 III-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 (12) and its physical mechanism.

Fig. 10 plots the incoming signal wave to the receiver, in which $d_0$ 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 wave distance increases a wavelength $\lambda$, the wave phase increases $2\pi$. Thus, the phase difference between adjacent elements is

$$\frac{2\pi d_a \sin \varphi}{\lambda}. \quad (23)$$

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

$$a_r(\varphi) = \frac{1}{\sqrt{N_r}} \left[1, e^{-j\frac{2\pi d_a \sin \varphi}{\lambda}}, \ldots, e^{-j\frac{2\pi (N_r-1) d_a \sin \varphi}{\lambda}} \right]^T, \quad (24)$$

where $k = 2\pi/\lambda$. Next, we study how to use the phased array to combine the received signals. Substituting (22) into (21), we aim to find the optimal combiner to maximize the resulting power, which is equivalent to

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

s.t. $|w(i)| = 1$, $i = 1, 2, \ldots, N_r$.

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

$$w = a_r(\varphi), \quad (25)$$

by which the resulting signal is the amplified one received at the first antenna. Thus, by using the combiner $a_r(\varphi)$, we aim to receive the narrow beam from the direction of $\varphi$. This process can be regarded as an inverse beamforming, also known as combining (24).

The array response vector in (24) represents the form of the signal coming from the direction $\varphi$. On the contrary, if the antenna elements transmit signals with the phase difference in (23), the wavefront moves in the direction of $\varphi$. To be exact, as shown in Fig. 11 assuming the transmitted signal at the first element is $c_t \in \mathbb{C}$, in the meantime, the transmitted signal at the $n$-th element should be $c_t \cdot e^{j \frac{2\pi (n-1)d_a \sin \varphi}{\lambda}}$. As such, the signal waves in the direction of $\varphi$ has the same phase and will add together to increase the radiation. 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

$$a(\varphi) = \frac{1}{\sqrt{N_a}} \left[1, e^{j\pi \sin \varphi}, \ldots, e^{j\pi (N_a-1) \sin \varphi} \right]^T. \quad (26)$$

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 \(\theta_1 = \theta_2 = \ldots = \theta_5\). With the near RF source, the AoA varies for each element, i.e., \(\theta_1 > \theta_2 > \ldots > \theta_5\). For this case, we need to do antenna testing and calibration to work out all these angles. One interesting question is when can we make the far-field assumption. In general, far-field is considered when the Rayleigh distance \(d\) is greater than
\[
d > \frac{2D^2}{\lambda},
\]
where \(D = (N_a - 1)d_a\) is the scale of the antenna array. That is to say, with the same scale of antenna array, the far-field distance 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 far-field distance 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 space holds \(d_a = \lambda/2\). As such, by substituting \(D = (N_a - 1)\lambda/2\), the condition \((27)\) can be rewritten as
\[
d > D(N_a - 1) = (N_a - 1)^2 \frac{\lambda}{2},
\]
which indicates that with the same number of antenna elements, the far-field distance for THz communication could be quite small. Consider two practical examples based on \((27)\) and \((28)\), respectively. (1) \(D = 10\) cm, \(f = 0.3\) THz \(\Rightarrow d = 20\) m. (2) \(N_a = 101\), \(f = 0.3\) THz \(\Rightarrow d = 5\) m. Note that \(D\) and \(N_a\) are defined for the one-dimension. As for URPA, the corresponding size and antenna numbers are 10 cm × 10 cm and 101 × 101, respectively. Based on the calculation for URPA, even using 2,000 antenna elements, the far-field assumption still holds when the Rayleigh distance is more than 1 meter.

In this subsection, we have briefly reviewed how to generate or combine a beam in a specific direction. Next, we introduce a feasible beamforming scheme, known as beam training, for establishing beam connections in THz UM-MIMO systems.

C. Beam Training

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
\[
P(2) : \{w^{\text{opt}}, f^{\text{opt}}\} = \arg \max \|w^H H f\|^2
\text{ s.t. } \|f\|^2 \leq 1, \|w\|^2 \leq 1.
\]
Provided that \(H\) is perfectly known at the transmitter and receiver, the optimal beamformer \(f^{\text{opt}}\) and the combiner \(w^{\text{opt}}\) can be obtained by leveraging the singular value decomposition (SVD) on \(H\). However, the conventional channel estimation methods tailored for microwave and mmWave communication may not apply to THz UM-MIMO systems since their pilot signals suffer from severe path loss and cannot be effectively detected by the receiver before efficient beamforming. Considering this fact, the transmitter and the receiver need to find \(f^{\text{opt}}\) and \(w^{\text{opt}}\) without the explicit channel state information. Since the UM-MIMO channels are comprised of multiple paths with high directivity and approximate orthogonality \([125]\). A near-optimal solution for the transmitter (resp. the receiver) is to form (resp. combine) a beam along the direction of paths, as shown in Fig. 13. By considering the LoS channel, the beam training problem is given by
\[
P(3) : \{w^{\text{opt}}, f^{\text{opt}}\} = \arg \max \|w^H H_{\text{LoS}} f\|^2
\text{ s.t. } H_{\text{LoS}} = \alpha_{11} a_{1} a_{1}^H, \quad f \in \mathcal{F}, \quad w \in \mathcal{W},
\]
where \(\mathcal{F}\) and \(\mathcal{W}\) are predefined codebooks for transmitter and receiver, respectively. In this case, beam steering is an optimal solution, i.e., \(w^{\text{opt}} = a_{\tau}\) and \(f^{\text{opt}} = a_{\tau}\). For multi-access scenarios, such as interference channel (IFC) and interferring broadcast channel (IBC) in single/multi-cell MIMO systems, interference exists in the digital baseband due to the non-orthogonal beam steering in analog domain. The interference can be eliminated by digital precoding schemes to further enhance spectral efficiency, e.g., by using the zero forcing (ZF) or block diagonalization (BD) technologies \([126]-[128]\).

As we mentioned in Section III-B, the array response vector act as a form of the narrow beam. To realize the beam steering in UM-MIMO systems, a common way is to predefined a codebook that includes many antenna response vectors representing the beams with different space angles \([125],[129],[130]\). By testing the narrow beam pairs (beamformer-combiner pairs) of the transmitter and the receiver, the system can obtain the optimal beam-steering solution, i.e., the pair with the strongest signal power. The search for the best narrow beam pair is often referred to as the beam training \([131]-[133]\).

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 channel estimation and precoding optimization are further needed before transmitting data \([134]-[136]\). Whereas, beam training focuses on seeking the strongest narrow beam pair and then directly transmits data without other channel information.

Next, we show some important features of the beam training by analyzing its performance on the ULA. In this simple case, the beam training is equivalent to find \(a_{\tau}(\varphi_\tau)\) and \(a_{\tau}(\varphi_\tau)\) by...
testing a codebook as
\[
\mathcal{F} = \{a_1(\varphi), a_2(\varphi), \ldots, a_N(\varphi)\},
\]
\[
\mathcal{W} = \{a_1(e^{j\varphi}), a_2(e^{j\varphi}), \ldots, a_N(e^{j\varphi})\},
\]
where \(N\) is the number of codewords at transmitter and receiver, respectively.

Despite \(\varphi_r \in [0, 2\pi]\) and \(\varphi_t \in [0, 2\pi]\), it suffices to consider the \(N\) narrow 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.

**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 \(\varphi\) is equivalent to that in direction of \(\pi - \varphi\), i.e.,
\[
a(\varphi) = a(\pi - \varphi).
\]

**Proof:** Following the convention, we define the angle perpendicular to the array plane as 0 and accordingly the front range is \([-\frac{\pi}{2}, \frac{\pi}{2}\]\). It is easy to verify \(30\) from the expression shown in \(26\) due to the fact \(\sin(\varphi) = \sin(\pi - \varphi)\).

Then, we define the beam coverage of the \(N_a\) narrow beam \(a(\varphi)\) used for beam training as
\[
\mathcal{CV}(a(\varphi)) = \{\psi : A(a(\varphi), \psi) \geq \rho\},
\]
where \(A(a(\varphi), \psi)\) is the normalized beam gain of \(a(\varphi)\) in the direction of \(\psi\), i.e., \(A(a(\varphi), \psi) = \|a(\varphi)^H a(\psi)\|\). \(\rho\) is the beam gain threshold. The normalized beam gain of \(A(a(\varphi), \psi)\) characterizes the radiation pattern of the beam \(a(\varphi)\) in magnitude, and can be regarded as the beam’s shape. Fig. 14 plots the radiation patterns of the narrow beam \(a(0)\) with elements \(N_a\) of 8, 16 and 64. It is observed that the beam has considerable gain in the beam direction, and the beam is sharper with the increase of the number of antenna elements. In addition, by plotting narrow beams in the polar coordinate plane, we show the shape of them in Fig. 15. As can be seen, the beam is narrower in the direction around 0 (perpendicular to the array) and is wider in the direction around \(\pm \frac{\pi}{2}\). Each beam actually shows that the symmetrical radiation patterns within the range \([-\frac{\pi}{2}, \frac{\pi}{2}\]\) and \([\frac{\pi}{2}, \frac{3\pi}{2}\]\). This observation also validates Lemma 1.

We should mention that the normalized beam gain in magnitude is also named as array factor in antenna theory \(96\), which can be rewritten as
\[
A(a(\varphi), \psi) = \frac{1}{N_a} \sum_{n=1}^{N_a} e^{jkd_a(n-1)}[\sin(\psi) - \sin(\varphi)]
\]
\[
= \frac{1}{N_a} e^{jkd_a m} \left( e^{jN_a kda m} - e^{-jN_a kda m} \right)
\]
\[
= \frac{1}{N_a} e^{j(N_a-1)kda m} \left( \sin((N_a kda m)/2) \right)
\]
\[
\sin((kda m)/2)
\]
where \(m = \sin(\psi) - \sin(\varphi)\). Substituting \(d_a = \lambda/2\), the normalized beam gain is given by
\[
A(a(\varphi), \psi) = \frac{\sin(N_a \frac{\pi}{2} (\sin \psi - \sin \varphi))}{N_a \sin(\frac{\pi}{2} (\sin \psi - \sin \varphi))}
\]
Define \(T(x) = |\sin(N_a \frac{\pi}{2} x)|/|N_a \sin(\frac{\pi}{2} x)|\) and we show the function of \(33\) in Fig. 16. It is observed that the gain can be maximally detected in the direction of \(\varphi\) and decreases in other directions. In this regard, \(\rho\) in \(31\) denotes a coverage
threshold, i.e., the direction range within which the normalized beam gain is no smaller than \( \rho \). The length of the range is called beamwidth. When \( \rho = \frac{1}{2} \), i.e., the beam power decreases 3dB, it is defined as the well-known HPBW. 

Let \( N_a \) and \( N \) be the number of antenna elements and the number of beams, respectively. Fig.[17] shows the cases of \( N = N_a \) beams and \( N = 2N_a \) beams covering the whole space. As shown in Fig.[17] the best case of the beam training is that the path angle exactly coincides with the angle of one beam center, and the worst case is that the path angle lies on the intersection of the beams. With \( N \) beams covering the whole space, the beam gain in worst case is given by

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

(34)

It is obvious that more beams in the codebook, better beam gain yields in statistics. However, with the implementation of UM-MIMO, the exhaustive searching of the optimal beam pair is quite time consuming. Considering both the transmitter and the receiver have \( N \) predefined narrow beams, the complexity of exhaustively searching the narrow beam pair is of

\[
T_{\text{exh}} = N^2. 
\]  

(35)

Compared to exhaustive searching, some training procedures are more appealing owing to their effectiveness with lower time complexity [116], [131], [137]–[140]. IEEE 802.11ad proposed an one-sided search algorithm [137], 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 [138] proposed a parallel beam search approach which uses \( N_{\text{RF}} \) RF chains at BS to transmit multiple beams simultaneously while all users exhaustively search the beams, which incurs the complexity of \( N^2/N_{\text{RF}} \). Apart from the above approaches, an appealing concept is known as the hierarchical beam training, which first searches wide beams and then searches narrow beams.

There are various schemes for realizing hierarchical training. The authors in [139] proposed to use fewer elements to transmit low-frequency signals for realizing wide beams and then use massive elements to achieve the fine search by high-frequency narrow beams. The authors in [116] 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 on-trend manner of hierarchical training is the M-tree search proposed in [131] and [140]. In the M-tree, there are \( S = \log_M N \) stages and the \( s \)-th stage has \( M^s \) beams. Let \( \omega_n \) denote the \( n \)-th beam in the \( s \)-th stage. Fig.[18] shows the beam coverage structure of a ternary-tree codebook, i.e., \( M = 3 \). As can be seen, each wide beam exactly covers three narrower beams in the next stage. Fig.[19] 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.[20]. In the one-side tree search [131], 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

\[
T_{\text{one}} = M \log_M N + M \log_M N = 2M \log_M N. 
\]  

(36)

In the both-side tree search [140], 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

\[
T_{\text{both}} = M^2 \log_M N. 
\]  

(37)

Notice from (36) and (37) that the complexity of the tree search is much less than the complexity of the exhaustive search. In
In general, there are three main fabrication technologies for antenna elements located in the focal region of the lens \[143\]. Array connected with transmission lines with variable lengths \[145\], \[146\]; 2) by dielectric materials with carefully designed inductive and capacitor structures \[149\], \[150\].

Instead, only a simple link selection is required, which makes the wireless communication system particularly referred to as the beamforming realized by using the special hardware that makes the wireless communication system more like an optical one \[141\]. 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 \[142\].

Fig. 21. Architecture of the training tailored hardware.

As shown in Fig. 22, the authors in \[151\] presented a prototype of the lens array that consists of four main components: 1) field programmable gate array (FPGA)-based digital signal processor (DSP) back-end and analog-to-digital conversions (ADCs), 2) RF chains, 3) beam selector, and 4) front-end lens array. It can be observed that the fundamental principle of the lens array can focus the incident signals with sufficiently separated AoAs to different antenna elements (or a subset of elements), and vice versa. In works \[144\], \[152\]–\[155\], the lens array has been shown to achieve significant performance gains as well as complexity reductions in UM-MIMO systems. Based on the lens array, the signal processing can be much simplified by treating the transmit/received signals on the virtual channels. Specifically, the conventional UM-MIMO system can be expressed as

\[
y = \mathbf{H} \mathbf{x} + \mathbf{n},
\]

where \( \mathbf{H} \in \mathbb{C}^{N_r \times N_t} \) and \( \mathbf{x} \in \mathbb{C}^{N_t} \) are the transmit and receive signals on the antenna elements, which have large dimensions, respectively. By using the lens array, the UM-MIMO system can be expressed as

\[
\tilde{y} = \mathbf{W}_{\text{lens}} \mathbf{H}_{\text{lens}} \tilde{x} + \tilde{n} = \tilde{H} \mathbf{x} + \tilde{n},
\]

where \( \tilde{y} \in \mathbb{C}^{N_r} \) and \( \tilde{x} \in \mathbb{C}^{N_t} \) are the transmit and receive signals on the lens array's elements, respectively. \( \mathbf{W}_{\text{lens}} \in \mathbb{C}^{N_r \times N_t} \) and \( \mathbf{F}_{\text{lens}} \in \mathbb{C}^{N_t \times N_t} \) are the fixed analog beamforming in the lens array architecture, which represents the signals' transformation from the lens to the antenna elements at the transmitter and receiver, respectively. \( \tilde{H} = \mathbf{W}_{\text{lens}} \mathbf{H}_{\text{lens}} \) is the virtual channel of the UM-MIMO systems based on lens array. Generally, the fixed analog beamforming matrices \( \mathbf{W}_{\text{lens}} \) and \( \mathbf{F}_{\text{lens}} \) can be expressed as the unitary discrete Fourier transform (DFT) matrices \[156\], whose columns correspond to orthogonal beams with different spatial angles. An \( N \times N \) DFT matrix \( \mathbf{U} \) is given by

D. Beamspace MIMO

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 \[141\]. 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 \[142\].

Fig. 22. An lens array prototype \[151\]: (a) block diagram of the transceiver; (b) block diagram of the RF chains; (c) structure of the lens array.

Addition, the binary tree, i.e., \( M = 2 \), has the same complexity for both the one-side and both-side searches.

Fig. 23. Beam selection in beamspace MIMO systems with five users.
U = \frac{1}{N} \times 
\begin{bmatrix}
1 & e^{j\pi \sin \varphi_1} & \cdots & e^{j\pi \sin \varphi_N} \\
e^{j\pi (N-1) \sin \varphi_1} & e^{j\pi (N-1) \sin \varphi_2} & \cdots & e^{j\pi (N-1) \sin \varphi_N}
\end{bmatrix}

(39)

It is worth pointing out that the virtual channel $\tilde{H}$ has a sparse property in the THz UM-MIMO systems due to the limited number of spatial paths. Thus, the transmit vector $\tilde{x}$ and receive vector $\tilde{y}$ has limited elements with large magnitude. In view of this, we can merely process the signals from the elements with large magnitude and reduce the dimension from $N_t/N_c$ to $N_t^{RF}/N_c^{RF}$, where $N_t^{RF}$ and $N_c^{RF}$ represent the number of RF chains of the transmitter and receiver respectively. This is referred to as the antenna selection or beam selection [141], [144], [157]. For ease of understanding, an illustration of the antenna selection in beamspace multi-user MIMO system is presented in Fig. 23. The less the number of users served, the less the number of antenna elements selected.

IV. WIDEBAND BEAMFORMING

In the previous section, we have illustrated the principle of single-frequency beamforming via phased array. In this section, we discuss two main effects, i.e., spatial-wideband effect and frequency-wideband effect, which need to be considered in the wideband beamforming by using phased array as they may significantly affect THz communications performance.

A. Spatial-Wideband Effect

Now, we consider a wideband communication with a very high symbol rate. When the plane wave arrives slanting 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 [158]–[160].

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.,

$$g(t) = \begin{cases} 1, & 0 \leq t \leq T_s \\ 0, & \text{otherwise} \end{cases}$$

(40)

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. 24 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

$$\hat{\tau}_m = \frac{(m-1)d_a \sin \varphi}{c}.$$  

(41)

The phase difference at the $m$-th antenna element is $2\pi \cdot \hat{\tau}_m \cdot f = \frac{2\pi (m-1)d_a \sin \varphi}{\lambda} f$. Thus, the equivalent baseband signal received by the $m$-th antenna element can be written as

$$y_m(t) = \alpha s \left( t - \frac{(m-1)d_a \sin \varphi}{c} \right) e^{-j2\pi (m-1)d_a \sin \varphi}.$$  

(42)

Let $\hat{\tau}_{\text{max}} = \frac{(N_a-1)d_a \sin \varphi}{c}$ 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 $\mathbf{H}$. 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 $N_a$ yields possibly larger $\hat{\tau}_{\text{max}}$, leading to the spatial-wideband effect determined by the array size and symbol rate.

A reasonable solution to the above problem is dividing the baseband into several sub-bands and concurrently processing each sub-band that has a lower symbol rate by the digital beamformer. For each sub-band, the equivalent symbol period is enlarged, which is beneficial to mitigate the spatial-wideband effect. On the other hand, we can use the true-time-delay (TTD)-based precoder/combiner for the wideband beamforming [161], which can also address the frequency-wideband effect, to be discussed in the next subsection.

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 by 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 of a phase-array codeword changes with the frequency of the signal, which is known as beam squint effect [162]–[164].

The beam squint effects take place at both the transmitter and the receiver. Without loss of generality, we analyze its effect in
insights for the beam squint effects as follow.

When \( \psi = 90^\circ \) incoming signals with AoA of \( \psi \) and \( f \) are carrier frequency. Let \( B \) represent the baseband bandwidth. Then, we have \( f \in [f_c - B, f_c + B] \) and \( \xi \in [1 - B/f_c, 1 + B/f_c] \). Commonly, the combiner is set based on the carrier frequency, i.e., \( A(\varphi) \equiv A(\varphi, f_c) \). Thus, the normalized beam gain of the combiner \( A(\varphi) \) in the direction of \( \psi \) at frequency \( f \) can be expressed as

\[
A_f (a(\varphi), \psi) = \begin{bmatrix} e^{-j2\pi f d_a (n-1) \sin \psi} \end{bmatrix}^T.
\]

(43)

Let \( \xi = \frac{f}{f_c} \) and \( f_c \) be the carrier frequency. Then, we have \( f \in [f_c - B, f_c + B] \) and \( \xi \in [1 - B/f_c, 1 + B/f_c] \). Commonly, the combiner is set based on the carrier frequency, i.e., \( a(\varphi) \equiv a(\varphi, f_c) \). Thus, the normalized beam gain of the combiner \( A(\varphi) \) in the direction of \( \psi \) at frequency \( f \) can be expressed as

\[
A_f (a(\varphi), \psi) = \begin{bmatrix} e^{-j2\pi f d_a (n-1) \sin \psi} \end{bmatrix}^T.
\]

(44)

Similar to the derivation in (32) and (33), (44) can be rewritten as

\[
A_f (a(\varphi), \psi) = \begin{bmatrix} \sin \frac{\xi}{d_a} (\xi \sin \psi - \sin \varphi) \\
\frac{1}{d_a} (\xi \sin \psi - \sin \varphi) \end{bmatrix}.
\]

(45)

When \( \xi = 1 \), i.e., \( f = f_c \), \( A_f (a(\varphi), \psi) \) is reduced into the narrowband form in (33). Consider a wideband incoming signals with AoA \( \psi = \pi/4 \), \( f_c = 0.14 \) THz, \( B = 10 \) GHz, \( N_a = 80 \). Fig. [25] 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 squints to some degree. It is worth mentioning that the beam squint direction is different between the transmitter and the receiver. To be exact, as shown in Fig. [25] the receive beam pattern with high frequency moves towards 90 degrees while that with low-frequency moves towards 0. However, the transmit beam pattern with high frequency moves towards 0 while the low-frequency beam moves towards 90 degrees.

Let us revisit the normalized beam gain given in (45). For incoming signals with AoA of \( \psi \) at frequency \( f \), the optimal combiner \( a(\varphi) \) satisfies

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

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

(46)

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

- When \( f = f_c \), there is no beam squint as \( \varphi = \psi \).
C. TTD Architectures

The TTD architecture can be used to address both spatial-wideband and frequency-wideband effects, as it compensates for the effects of space and frequency by imposing a true time delay on the signal at each antenna element. The TTD requirement can easily be realized in the digital domain but suffers from high hardware complexity and cost in analog domain as there are many RF chains and the dynamic range requirement is extremely high. To achieve the TTD in the analog domain, we can prolong the propagation length or slow the wave velocity of the electromagnetic signals in the precoder/combiner. There has been little success in reducing the wave velocity and the former one is mainly adopted in existing works [166–168]. Some emerging THz TTDs have been reported in [169–172].

There are four feasible TTD architectures in wideband beamforming, as presented in Fig. 26 in Fig. 26 (a) and (b), full-connected TTD (FC-TTD) architecture and partially-connected TTD (PC-TTD) architecture replace the phase shift matrix in (18) and (19) 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 [62], 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. 26 (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 [164]. Similarly, the authors in [173] proposed to combine the TTD devices and the conventional phase shifts to jointly design wideband beamforming with reduced implementation cost compared to the TTD arrays. The infinite-precision TTD is impractical for hardware implementation. As a solution, a switching network is designed to flexibly select the appropriate time delay from a set of fixed TTDs, as shown in Fig. 26 (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 [64], [174].

V. IRS-ASSISTED JOINT BEAMFORMING

Despite UM-MIMO technologies are able to offer great beamforming gains, its energy efficiency decreases with the increase of the number of antenna elements [175]. This is because the total energy consumption increases linearly w.r.t. the number of active components, whereas the data rates only grow logarithmically [176]. In this subsection, we consider an emerging wireless technology, i.e., IRS, which is viewed as an appealing complement for MIMO systems as it can significantly improve the spectral efficiency with much reduced energy consumption [177–181].

IRS is a metasurface consisting of a large set of tiny elements (i.e., controllable reflecting elements), each being able to passively 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/countered in intended/unintended directions [182]. 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 [183]. 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 multi-input single-output (MISO) IRS-assisted system [184]–[188]. With different optimization targets, transmit power minimization [184], weighted sum-rate maximization [185], energy efficiency maximization [186], multicast rate maximization [187], and latency minimization [188] have been investigated.

A. IRS-Assisted UM-MIMO Systems

Consider a point-to-point IRS-assisted UM-MIMO communication system as depicted in Fig. 27, 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 $\mathbf{s} \in \mathbb{C}^{N_s \times 1}$ to the user and the IRS simultaneously. Let $\mathbf{H}_{\text{LoS}} \in \mathbb{C}^{N_r \times N_t}$, $\mathbf{M} \in \mathbb{C}^{N_{IRS} \times N_t}$, and $\mathbf{N} \in \mathbb{C}^{N_r \times N_{IRS}}$ 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 $\mathbf{\Theta} = \text{diag}(\beta e^{j\theta_1}, \beta e^{j\theta_2}, \ldots, \beta e^{j\theta_{N_{IRS}}}) \in \mathbb{C}^{N_{IRS} \times N_{IRS}}$ 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

$$
\mathbf{y} = \sqrt{\frac{P}{N_s}} \mathbf{N} \mathbf{M} \mathbf{F} \mathbf{s} + \sqrt{P} \mathbf{H}_{\text{LoS}} \mathbf{F} \mathbf{s} + \mathbf{n},
$$

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

\[
P(4) : \max_{\mathbf{F}, \mathbf{v}} \log_2 \det \left( \mathbf{I}_{N_s} + \frac{P}{\sigma^2} \mathbf{H}_{\text{eff}} \mathbf{F}^H \mathbf{H}_{\text{eff}} \right)
\]

s.t. \( \mathbf{H}_{\text{eff}} = \mathbf{N} \Theta \mathbf{M} + \mathbf{H}_{\text{LoS}} \),

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

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

where \( \mathbf{v}(i) \) denotes the \( i \)-th entry of \( \mathbf{v} \). For UM-MIMO systems, \( P(4) \) 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 \( \mathbf{v} \). To solve this problem, the authors in [189] proposed a sum-path-gain maximization approach to reach a suboptimal solution. Then, the authors in [190] 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 [191] 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 [192]–[196] and the beamforming optimization problems [197]–[202] 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, as their estimation approaches can hardly establish the beam alignment and their beamforming optimization could result in extremely high implementation complexity in the case of the tremendous number of antenna elements. Thus, it is of vital importance to consider the beam training strategies for IRS-assisted systems [203]–[206]. 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 [203] 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 [204] 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 [205] and [206], 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. 28, the IRS-assisted system consists of six path angles \( \phi_{B,H}, \phi_{U,H}, \phi_{B,M}, \phi_{R,M}, \phi_{R,N}, \) and \( \phi_{U,N} \). The 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. 29, we use a solid red line to represent the transmit beams, a solid-broken blue 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 of the passive terminal (IRS), the codewords are functions of the angle, i.e., \( \{ \Theta_{\text{ret}}(\varphi_{in}^m) \}_{n=1}^N \). If the AoA of incoming signals is \( \varphi_{in}^m \), the codeword \( \Theta_{\text{ret}}(\varphi_{in}^m) \) ensures that the AoD of reflected signals is the back direction \( \varphi_{in} + \pi \), i.e., \( \Theta_{\text{ret}}(\varphi_{in}^m)a(\varphi_{in}^m) = a(\varphi_{in} + \pi) \). The codewords of the directions mode are functions of two angles, i.e., \( \{ \Theta_{\text{dir}}(\varphi_{in}^m, \varphi_p) \}_{n=1}^N \). If the AoA of incoming signals is \( \varphi_{in}^m \), the codeword \( \Theta_{\text{dir}}(\varphi_{in}^m, \varphi_{in}^m) \) ensures that the
suffers from the following main drawbacks.

and the user, we obtain

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

determining the time slot with the strongest receiving signal.

ϕ

By switching the operation of the BS and the user, we obtain

ϕ

 omitted the codeword design. Interested readers can learn about

the details in [206].

In the following content, we first present the primary idea

of the IRS-assisted joint beam training, albeit with some sh ort-

comings, to draw some basic insights. We divide the overall

procedure into three phases to achieve different groups of

measurements as illustrated in Fig. 30.

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 ϕU,H and ϕR,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., Θret(ϕ

), which are predefined in time slots and 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

ϕR,M and ϕR,N.

Phase 3: With the obtained ϕR,M and ϕR,N, we fix IRS

to optimally bridge the BS-IRS-user link by direction mode, i.e., Θdir(ϕR,M,ϕR,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 ϕU,N and ϕR,B,M.

Based on the above three phases, we can find all the path

angles in IRS-assisted systems [205]. However, this strategy

suffers from the following main drawbacks.

• Omni-beam may not be effectively detected in THz com-

munication.

• Concurrently transmitting and receiving beams result in interference.

AoD of reflected signals is ϕ\text{out}_m, i.e., Θ_{\text{dir}}(ϕ_k,ϕ\text{out}_m)\text{a}(ϕ_k) = a(ϕ\text{out}_m). Here, we only discuss the effects of these modes and

omit the codeword design. Interested readers can learn about

the details in [206].

In the following content, we first present the primary idea

of the IRS-assisted joint beam training, albeit with some sh ort-

comings, to draw some basic insights. We divide the overall

procedure into three phases to achieve different groups of

measurements as illustrated in Fig. 30.

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 ϕU,H and ϕR,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., Θret(ϕ\text{in}_n), which are predefined in time slots and 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

ϕR,M and ϕR,N.

Phase 3: With the obtained ϕR,M and ϕR,N, we fix IRS

to optimally bridge the BS-IRS-user link by direction mode, i.e., Θ_{\text{dir}}(ϕR,M,ϕR,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 ϕU,N and ϕR,B,M.

Based on the above three phases, we can find all the path

angles in IRS-assisted systems [205]. However, this strategy

suffers from the following main drawbacks.

• Omni-beam may not be effectively detected in THz com-

munication.

• Concurrently transmitting and receiving beams result in interference.

Fig. 30. Primary idea of the IRS-assisted joint beam training.

In the following content, we first present the primary idea

of the IRS-assisted joint beam training, albeit with some sh ort-

comings, to draw some basic insights. We divide the overall

procedure into three phases to achieve different groups of

measurements as illustrated in Fig. 30.

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 ϕU,H and ϕR,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., Θret(ϕ\text{in}_n), which are predefined in time slots and 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

ϕR,M and ϕR,N.

Phase 3: With the obtained ϕR,M and ϕR,N, we fix IRS

to optimally bridge the BS-IRS-user link by direction mode, i.e., Θ_{\text{dir}}(ϕR,M,ϕR,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 ϕU,N and ϕR,B,M.

Based on the above three phases, we can find all the path

angles in IRS-assisted systems [205]. However, this strategy

suffers from the following main drawbacks.

• Omni-beam may not be effectively detected in THz com-

munication.

• Concurrently transmitting and receiving beams result in interference.

Fig. 31. Practical beam training procedure for IRS-assisted systems.

• Sweeping the narrow beams incurs high complexity.

Focusing on these issues, we now extend this strategy to a more practical procedure. In this procedure, assuming that the path

angles ϕR,M and ϕR,N at IRS has only N value points, we need
to judiciously predefine 2N + 1 codewords for IRS (the details

of codeword design is referred to [205]), i.e., \{Θ_n\}_{n=1}^{2N + 1}, where the optimal phase shifts at IRS is covered by the codewords, i.e., Θ_{\text{dir}}(ϕR,M,ϕR,N) ∈ \{Θ_n\}_{n=1}^{2N + 1}. Combining with the hierarchical search introduced in Section III-C, we show the diagram of the practical procedure in Fig. 31.

Phase 1: In Phase 1, we aim to obtain the optimal codeword

for IRS. To achieve the beam alignment, we first test 3 × 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 \{Θ_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 Θ_{\text{dir}}(ϕR,M,ϕR,N). Thus, the

user can utilize the pulse slots to identify this optimal codeword

for IRS.

Phase 2: In Phase 2, we turn off the IRS and aim to obtain

ϕB,H and ϕU,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-
Based on 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 beam training has $N^2 + N^4$ tests in IRS-assisted systems, whereas this training procedure has only $18N + 12\log_3N - 3$ tests [206]. Now let us 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 sides of 5 m. The number 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. 32 shows the performance of different schemes in THz LoS-blockage case, where $H_{\text{LoS}} = 0$. 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 [207]. 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. 32 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 [208]. 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 [209]. 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) [210].

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.
can be regarded as a compression process from complexity channel estimation and precoding. The beam alignment merely to the digital-port channel matrix domain, the hybrid beamforming architecture applies the UM-digital ports. By operating the beam alignment in the analog video equivalent beam gain with fewer RF chains, i.e., less of partially-connected hybrid beamforming architecture can provide two DoF by using different antenna polarization modes.

B. THz UM-MIMO Beamforming Strategy

The conventional multi-user MIMO channel, denoted as $H_{\text{ant}}$, is defined between the transmit and receiver antenna ports. In the THz UM-MIMO systems, a large-scale antenna array needs to be employed at the transceivers for overcoming the severe propagation loss, which results in a large-dimension matrix $H_{\text{ant}}$. To leverage the fully DoF of THz UM-MIMO channel, accurate channel estimation for $H_{\text{ant}}$ requires extremely high-complexity array architecture and signal processor, which is not cost-efficient for THz UM-MIMO systems. As an effective solution, THz UM-MIMO systems can adopt a partial-connected hybrid beamforming architecture, which is also referred to as the array-of-subarray (AoSA) architecture, to reduce both hardware and algorithm overheads.

Specifically, compared to the digital beamforming, the partially-connected hybrid beamforming architecture can provide equivalent beam gain with fewer RF chains, i.e., less of digital ports. By operating the beam alignment in the analog domain, the hybrid beamforming architecture applies the UM-MIMO algorithms merely to the digital-port channel matrix $H_{\text{dig}}$ with a much smaller dimension, facilitating a low complexity channel estimation and precoding. The beam alignment can be regarded as a compression process from $H_{\text{ant}}$ to $H_{\text{dig}}$. By using a training codebook without complex calculations, this process can find the LoS component of the UM-MIMO channel. Although this strategy sacrifices some channel DoFs, its performance loss is not notable since $H_{\text{ant}}$ is LoS dominant and is sparse. For realizing the multiplexing, the LoS link can provide two DoF by using different antenna polarization modes.

C. Linear Beamforming Algorithms

The main goal of MIMO beamforming is to convey independent data streams to the served users and attain spatial multiplexing gain. It was shown that Costa’s dirty-paper precoding actually achieves the capacity region of the Gaussian BC multi-user MIMO channel. Achieving the capacity comes at the expense of non-linear processing. On the other hand, linear beamforming algorithms are more practical and well-explored in these years, which can be classified into 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). In the multiple-user MIMO scenarios, block-diagonalization (BD) 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) [217], eigen-based ZF (EZF) [218], coordinated BD (CBD) [219], regularized BD (RBD) [220], QR-Regularized BD (QR-RBD) [221], complex lattice reduction RBD (CLR-RBD) [222].

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 $1, 2, ..., k-1$, 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.

Fig. 34. A THz UM-MIMO system with partially-connected hybrid beamforming architectures.
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) [211]. 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) [224], minimum total mean square error criterion (T-MMSE) [225], and weighted sum minimum mean square error (WMMSE) [226].

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 $\mathbf{H}_{\text{dig}}$ 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 [216], EZF [218], QR-RBD [221], PMSE [224], and T-MMSE [225], 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. 35. 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 20 dB. 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. 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 [234], [235] 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 antennas 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.

![Fig. 35. The performance of different linear beamforming algorithms on THz UM-MIMO digital-port channel](image-url)
Due to the relatively mature integration capabilities of silicon-based, many small-scale phased arrays based on various silicon-based processes have been reported [246]–[253]. An example is given in [254], which shows a 4 × 4 URPA using 130 nm SiGe bipolar-CMOS (BiCMOS) technology at 320 GHz. As shown in Fig. 36 (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 [171] propose a 280 GHz × 2 × 2 chip-scale dielectric resonator antenna array. As shown in Fig. 36 (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 [172]. In [255], the proposed photoelectric phase shifter can control a 300 GHz beam scanning within 50 degree. Recently, the optical TTD phase shifters are also employed to offer stable time delay for wideband communications [172]. Fig. 36 (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 × 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., as 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. 36 (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. [268] envisages the use of graphene to implement UM-MIMO plasmonic nano-antenna arrays, which can implement a 1024 × 1024 UM-MIMO system at 1THz with arrays that occupy just 1 mm². Liquid crystal and graphene also show application potential in reconfigurable reflectarrays [269]–[272]. An example is given in [269], which gives the theoretical analysis of the LC reflectarray at 0.67 THz. As shown in Fig. 36 (e), it consists of a 24 element linear array, and each element is composed of 50 rows and 2 columns of unit cells with metal-insulator metal-resonator structure. In [271], the graphene-based reconfigurable metasurface is designed to achieve beam control at 2 THz. Authors in [272] experimentally demonstrate a 0.98 THz graphene reflectarray metasurface that can achieve THz beam steering. As shown in Fig. 36 (f), the reflectarray is...
mounted on a printed circuit board (PCB) and wire-bonded, where the substrate is comprising a reflective conductive ground plane and a dielectric spacer. It is worth pointing out that, so far, most new materials-based THz antenna arrays are still in the stage of theoretical design and analysis. The establishment of a complete array architecture for true THz frequencies with dynamic beamforming is not fulfilled. To sum up, Table VII shows the reported THz antenna arrays with dynamic beam scanning capability.

To sum up, some promising fabrication techniques for implementing UM-MIMO antenna arrays have been developed in the THz range, and future research may focus on seeking potential solutions for better beam flexibility as well as a larger array size, e.g., i) improving the dynamic beamforming capabilities of the THz arrays, including adjustment accuracy and scanning range, ii) increasing the size of the antenna array and pushing it to the level of thousands of elements, iii) and mutual coupling effects caused by large-scale integration.

### VIII. Emerging Applications

In this section, we introduce some applications for THz communication via beamforming technologies. Specifically, we enumerate six major scenarios, also shown in Fig. 37 as well as briefly illuminate 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 [278]. Recently, research progress has shown the prospect of using THz communication to build satellite networks, thanks to its high bandwidth and ignorable molecular absorption in this scenario. While retaining 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 [279].

**B. Vehicular Connectivity**

To realize the concept of the intelligent transportation system, vehicle communication network has been extensively studied, which roughly includes three types of connections: vehicle to vehicle (V2V), vehicle to infrastructure (V2I), and vehicle to anything (V2X) [43], [280]. 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 of less than 10 ms and reliability of 99.999%. For the era of

| Freq (Hz) | Size | Process | Beam scan | Gain | Antenna type | Reference |
|----------|------|---------|-----------|------|--------------|-----------|
| 280 G    | 4 × 4 arrays | 45 nm SOI CMOS | 80°/80° 1 | 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° 1 | - | 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° 1 | 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 | [171] |
| 300 G    | 1 × 4 arrays | photonic | 90° | 10.6 dBi | bow-tie antenna | [172] |
| 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 | [273] |
| 1.3 T    | 25448 elements | graphene | - | 29.3 dBi | reflectarray | [256] |
| 220 – 320 G 8 × 8 arrays | brass sheets | 50°/45° 1 | 17 dBi | frequency scanning | [274] |
| 100 G    | 54 × 52 cells | liquid crystals | 55° | 15 dBi | reflectarray | [275] |
| 345 G    | - | liquid crystals | 20° | 35 dBi | reflectarray | [276] |
| 100 G    | - | VO2 | 44°/44° | - | metasurfaces | [237] |
| 115 G    | 39 × 39 cells | liquid crystals | 20° | 16.55 dBi | reflectarray | [277] |

1 In both azimuth and elevation
wireless interconnected smart cars, THz communication will undoubtedly provide ultra-capacity, low-latency, high-mobility, and high-reliability for safe autonomous driving and intelligent transportation system (ITS) [53], [281]. It has been shown that THz wireless communication can support high-speed mobile railway application scenarios through dynamic beam tracking technology [282]. The channel propagation characteristics of the 300 GHz carrier in the V2I scenario have been explored in [283], which can be used to support the link-level and system-level design for future THz vehicular communications.

C. Wireless Local Area Networks

The ability of THz communication of achieving short-distance high-speed coverage in indoor scenarios can be resorted to enhance the WLAN. Considering that the transmission capacity of optical fiber links is much higher than that of WLAN, using THz wireless connects in the WLAN can enable seamless ultra-high-speed connections between high-speed wired networks and user mobile devices. Effective 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 [36]. 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 maximize the capabilities of T-WLAN [284]. 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 [285].

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 have high complexity, power consumption, maintenance costs, and space occupied by large cables. [286] 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 [287], [288]. Some researchers also point out that combining wireless communications with existing data center architectures can help reduce costs [289]. There have been some investigations on the THz channel modeling for wireless channel environment in data center [290]–[292].

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 [285], [293]. To be exact, the directional beams result in fewer scattering components and space sparsity, which brings a physical inability for any node located not in the beam direction to eavesdrop on the information [285]. In the meanwhile, As the transmit and receive THz beams are aligned within specific angle pairs, interference from the other angles can be effectively reduced.

Despite the path loss and directivity providing physical layer security, information security is still imperfect in the vicinity of the THz beam propagation path. Note that it is generally believed that the eavesdropper’s antenna must be located on the signal propagation path to monitor the signal. Since the THz...
beam is quite narrow, eavesdropping is not feasible. However, reveals 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 unexpected eavesdropper positions, unless counter-measures are used. To fix this defect, controls the coverage distance of the beam to improve the concealment of THz communication by controlling carrier frequencies, power allocation, and rate distribution on each sub-band. At the same time, it has been validated that integrating IRS to THz communication can further enhance the physical-layer security in the basic wiretap channel.

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 predication of complex multi-core wiring. 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.

IX. 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 [11A], 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. To achieve reliable channel modeling, channel measurement is an important means to verify and improve the 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, the measurements for about 1 THz is still limited.

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 due 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, materials such as SiGe, gallium nitride (GaN), indium phosphide (InP), quantum cascade laser (QCL), and graphene have been tried to achieve efficient generation and effective detection of THz waves. Nevertheless, there are still many challenges in hardware devices. 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. Graphene can be used to design compact THz transceivers due to its high conductivity and support the propagation of THz surface plasmon polaritons (SPP) waves. However, graphene is far less mature than the above-mentioned technologies due to the lack of research.

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, which includes the analysis of beamforming design under the influence of nonlinear distortion inflicted by low-resolution ADCs and quantized phase shifters.

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 alternative to conventional antennas, graphene-based large-scale antenna arrays are expected to overcome the above issues. The major challenge for graphene antennas is to characterize the interaction and coupling effects among adjacent elements. Current works of graphene-based antennas still focus on theoretical analysis but lack experimental exploration. To support the wideband THz communication, the design of wideband antennas is also worthy of further study.

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

The THz frequency band is recently becoming a new hot spot in the wireless communication community. Table VIII summarizes the features that distinguish sub-6 GHz, mmWave, and THz from the perspective of this review. In this paper, a primary tutorial on the beamforming technologies for the THz system was given. First, the model of the THz UM-MIMO systems was 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 and beamspace MIMO. Then, we proceeded to address the on-trend THz beamforming topics, i.e., wideband beamforming and IRS-assisted joint beamforming. For the former, we discussed the spatial-wideband and frequency-wideband effects, along with the feasible solutions. For the latter, we studied the model of IRS-assisted systems and provided a cooperative beam training scheme. Next, we provided a strategy for a multi-user UM-MIMO THz system and evaluated different linear beamforming algorithms for it. 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 addressed, which will be a crucial part of the future research directions.

F. Mobility Versus Blockage

The THz wave has limited diffraction and the LoS path is extremely easy to be blocked by obstacles. 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. Thus, THz communications need to expand coverage to support mobile access. On one hand, 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. With the exception of increasing the APs, the IRS-assisted solution supports the reconstruction of the wireless environment of THz propagation. Some key concerns include how to locate the IRS to enhance the signal coverage and how to realize the optimal routine. These concerns need to be fully addressed before practical systems can be deployed. Besides, in high-mobility outdoor scenarios, the beam misalignment caused by Doppler expansion needs to be seriously considered.

X. CONCLUSION

The THz frequency band is recently becoming a new hot spot in the wireless communication community. Table VIII summarizes the features that distinguish sub-6 GHz, mmWave, and THz from the perspective of this review. In this paper, a primary tutorial on the beamforming technologies for the THz system was given. First, the model of the THz UM-MIMO systems was 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 and beamspace MIMO. Then, we proceeded to address the on-trend THz beamforming topics, i.e., wideband beamforming and IRS-assisted joint beamforming. For the former, we discussed the spatial-wideband and frequency-wideband effects, along with the feasible solutions. For the latter, we studied the model of IRS-assisted systems and provided a cooperative beam training scheme. Next, we provided a strategy for a multi-user UM-MIMO THz system and evaluated different linear beamforming algorithms for it. 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 addressed, which will be a crucial part of the future research directions.
[305] J. Chen, “Hybrid beamforming with discrete phase shifters for millimeterwave massive MIMO systems,” IEEE Trans. Veh. Technol., vol. 66, no. 8, pp. 7604-7608, Aug. 2017.

[306] Y. Lin, “On the quantization of phase shifters for hybrid precoding systems,” IEEE Trans. Signal Process., vol. 65, no. 9, pp. 2237-2246, May 2017.

[307] U. Nissanov Nissan and G. Singh, “Terahertz Antenna for 5G Cellular Communication Systems: A Holistic Review,” in 2019 IEEE Int. Conf. Microwaves, Antennas, Commun. and Elect. Syst. (COMCAS), Tel-Aviv, Israel, 2019, pp. 1-6.

[308] 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.

[309] 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.

[310] 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.

[311] 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.

[312] M. T. Barros et al., “Integrated Terahertz Communication With Reflectors for 5G Small-Cell Networks,” IEEE Trans. Veh. Technol., vol. 66, no. 7, pp. 5647-5657, July 2017.

[313] Z. 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.

[314] 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.

[315] 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.