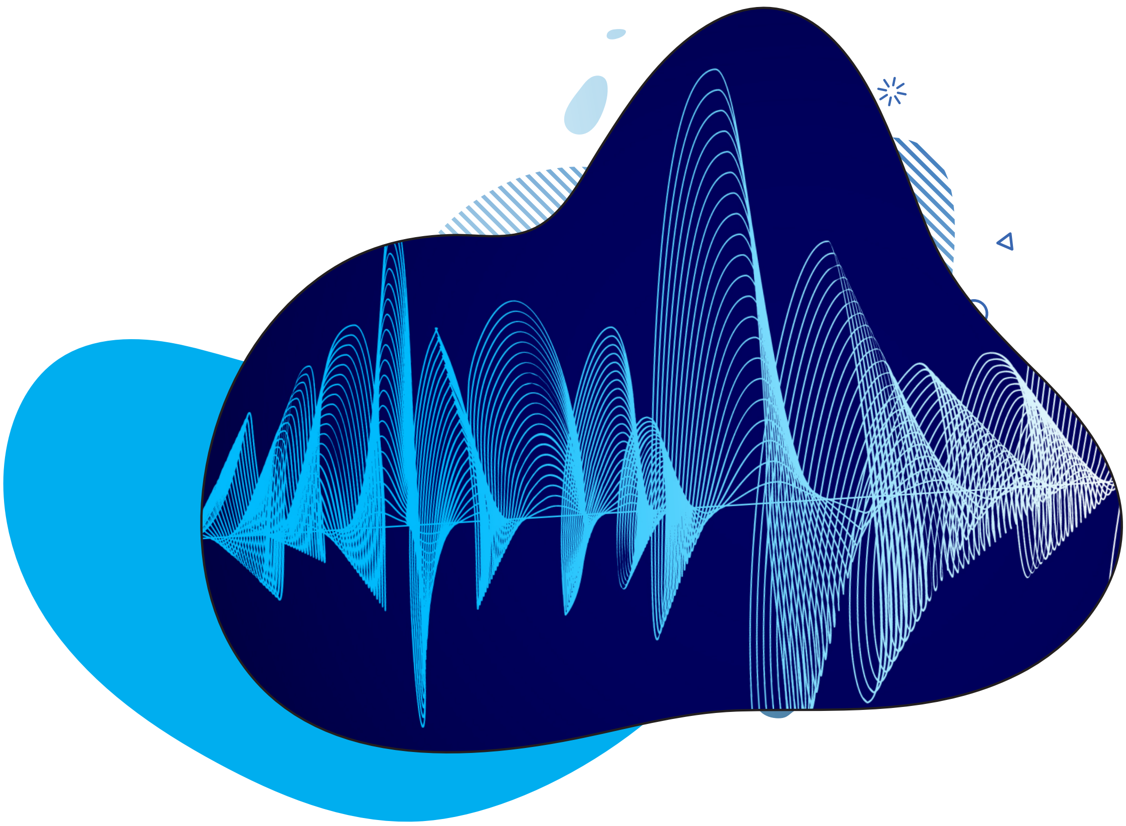
ITU Journal
Future and evolving technologies

FREE | FAST | FOR ALL

Special issue

Terahertz communications



Volume 2 (2021), Issue 7

ITU Journal
Future and evolving technologies

Volume 2 (2021), Issue 7

Terahertz communications

ITU Journal
Future and evolving technologies

FREE I FAST I FOR ALL

The ITU Journal on Future and Evolving Technologies (ITU J-FET) is an international journal providing complete coverage of all communications and networking paradigms, free of charge for both readers and authors.

The ITU Journal considers yet-to-be-published papers addressing fundamental and applied research. It shares new techniques and concepts, analyses and tutorials, and learnings from experiments and physical and simulated test beds. It also discusses the implications of the latest research results for policy and regulation, legal frameworks, and the economy and society. This publication builds bridges between disciplines, connects theory with application, and stimulates international dialogue. Its interdisciplinary approach reflects ITU's comprehensive field of interest and explores the convergence of ICT with other disciplines.

The ITU Journal welcomes submissions at any time, on any topic within its scope.

Publication rights

SUGGESTED CITATION: ITU Journal on Future and Evolving Technologies, Volume 2 (2021), Issue 7.

ADDITIONAL INFORMATION

Please visit the ITU J-FET website at
https://www.itu.int/en/journal/j-fet/Pages/default.aspx

Inquiries should be addressed to
Alessia Magliarditi at: journal@itu.int

EDITORIAL BOARD

TABLE OF CONTENTS

LIST OF ABSTRACTS

Terahertz band communications as a new frontier for drone networks

Pages 1–19

Akhtar Saeed, Ozgur Gurbuz, Mustafa Alper Akkaş, Ahmet Ozan Bicen

Terahertz band (0.1-10 THz) communications is one of the candidates for 6G systems due to intrinsic massive bandwidth and data rate support. Having demonstrated the significant potential of THz band at various atmospheric altitudes, in this article, we discuss the prospects of THz communications for drone networks, more specifically, for Drone Sensor Networks (DSNs). For 6G non-terrestrial communication scenarios, drones will not only serve as on-demand base-stations, as supporting alternatives or backhauls for the terrestrial base stations, but they will also provide seamless connectivity for distributed monitoring and surveillance applications, which require an ultra-reliable low latency service for carrying multimedia data. THz band sensing will also provide additional sensing capabilities from the sky to THz-enabled DSNs. Presenting this vision, in this paper, we first discuss possible use cases of THz-enabled drone networks considering communication, sensing and localization aspects. Then, for revealing the capacity potential of THz-enabled drone networks, we provide motivating channel capacity results for communication of drones at different altitudes, under ideal channel conditions with no fading and realistic channel with beam misalignment and multipath fading. We further present major challenges pertaining to employing the THz band for DSNs, addressing physical layer issues, followed with spectrum and interference management, medium access control and higher layers and security, while reviewing some prominent solutions. Finally, we highlight future research directions with Artificial Intelligence (AI)/Machine Learning (ML)-based approaches and mobile edge computing.

View Article

M-ary quadrature amplitude modulation order optimization for terahertz wireless communications over dispersive channels

Pages 21–30

Karl Strecker, Sabit Ekin, John O'Hara

Highly accurate atmospheric models, based on molecular resonance information contained within the HITRAN database, were used to simulate the propagation of high capacity single-carrier quadrature amplitude modulated signals through the atmosphere for various modulation orders. For high-bandwidth signals such as those considered in this work, group velocity dispersion caused by atmospheric gases distorts the modulated waveform, which may produce bit errors. This leads to stricter Signal-To-Noise Ratio requirements for error-free operation, and this effect is more pronounced in high-order modulation schemes. At the same time, high-order modulation schemes are more spectrally efficient, which reduces the bandwidth required to maintain a given data rate, and thus reduces the total group velocity dispersion in the link, resulting in less distortion and better performance. Our work with *M*-ary quadrature amplitude modulated signals shows that optimal selection of modulation order can minimize these conflicting effects, resulting in decreased error rate, and reducing the performance requirements placed on any equalizers, other dispersion-compensating technologies, or signal processing hardware.

View Article

Terahertz wireless communications in space

Pages 31–38
Meltem Civas, Ozgur B. Akan

The New Space Era has increased communication traffic in space by new space missions led by public space agencies and private companies. Mars colonization is also targeted by crewed missions in the near future. Due to increasing space traffic near Earth and Mars, the bandwidth is getting congested. Moreover, the downlink performance of the current missions is not satisfactory in terms of delay and data rate. Therefore, to meet the increasing demand in space links, Terahertz band (0.1–10 THz) wireless communications are proposed in this study. In line with this, we discuss the major challenges that the realization of THz band space links pose and possible solutions. Moreover, we simulate Mars-space THz links for the case of a clear Mars atmosphere, and a heavy dust storm to show that even in the worst conditions, a large bandwidth is available for Mars communication traffic.
View Article

Fundamental limits of high-efficiency silicon and compound semiconductor power amplifiers in 100–300 GHz bands

Pages 39–50
James F. Buckwalter, Mark J. W. Rodwell, Kang Ning, Ahmed Ahmed, Andrea Arias-Purdue, Jeff Chien, Everett O'Malley, Eythan Lam

This paper reviews the requirements for future digital arrays in terms of power amplifier requirements for output power and efficiency and the device technologies that will realize future energy-efficient communication and sensing electronics for the upper millimeter-wave bands (100–300 GHz). Fundamental device technologies are reviewed to compare the needs for compound semiconductors and silicon processes. Power amplifier circuit design above 100 GHz is reviewed based on load line and matching element losses. We present recently presented class-A and class-B PAs based on a InP HBT process that have demonstrated record efficiency and power around 140 GHz while discussing circuit techniques that can be applied in a variety of integrated circuits.
View Article

Self-configuring asynchronous sleeping in heterogeneous networks

Pages 51–62

Ali Medlej, Eugen Dedu, Kamal Beydoun, Dominique Dhoutaut

Nowadays, the heterogeneous wireless nano-network topology becomes a need for diverse applications based on heterogeneous networks composed of regions of different node densities. In Wireless Nano-networks (WNNs), nodes are of nano-metric size and can be potentially dense in terms of neighbouring nodes. Nano-nodes have limited resources in terms of processing, energy and memory capabilities. In nano-network(s), even in a communication range limited to tens of centimeters, thousands of neighbours can be found. We proposed a fine-grained duty-cycling method (sleeping mechanism), appropriate to nanonodes, which aims to reduce the number of receptions seen by a node during data packet routing. The present study reveals the usefulness of implementing the sleeping mechanism in heterogeneous networks, as well as configuring a dynamic awaken duration for nodes based on a density estimation algorithm. We also proposed an algorithm that helps in increasing the reliability of the packet received by the destination node.

View Article

Hierarchical beam alignment in SU-MIMO terahertz communications

Pages 63–80

Yifei Wu, Johannes Koch, Martin Vossiek, Wolfgang Gerstacker

Single-Carrier Frequency Division Multiple Access (SC-FDMA) is a promising technique for high data rate indoor Terahertz (THz) communications in future beyond 5G systems. In an indoor propagation scenario, the Line-Of-Sight (LOS) component may be blocked by the obstacles. Thus, efficient THz SC-FDMA communications require a fast and reliable Beam Alignment (BA) method for both LOS and Non-Line-Of-Sight (NLOS) scenarios. In this paper, we first adopt the hierarchical discrete Fourier transform codebook for LOS BA, and introduce the hierarchical k-means codebook for NLOS BA to improve the beamforming gain. Simulation results illustrate that the hierarchical DFT codebook and the hierarchical k-means codebook can achieve the beamforming gain close to that of the maximum ratio transmission in LOS and NLOS cases, respectively. Based on these two codebooks, we propose a Multi-Armed Bandit (MAB) algorithm named Hierarchical Beam Alignment (HBA) for single-user SC-FDMA THz systems to reduce the BA latency. HBA utilizes a hierarchical structure in the adopted codebook and prior knowledge regarding the noise power to speed up the BA process. Both theoretical analysis and simulation results indicate that the proposed BA method converges to the optimal beam with high probability for both the hierarchical DFT codebook and the hierarchical k-means codebook in the LOS and NLOS scenarios, respectively. The latency introduced by HBA is significantly lower when compared to an exhaustive search method and other MAB-based methods.

View Article

TERAHERTZ BAND COMMUNICATIONS AS A NEW FRONTIER FOR DRONE NETWORKS

Akhtar Saeed[1], Ozgur Gurbuz[1], Mustafa Alper Akkaş[2], Ahmet Ozan Bicen[3]
[1]Electronics Engineering Program, Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul, 34956, Turkey, [2]Department of Computer Engineering, Bolu Abant Izzet Baysal University, Bolu, 14280, Turkey, [3]Independent Contributor

NOTE: Corresponding author: Akhtar Saeed, akhtarsaeed@sabanciuniv.edu

Abstract – *Terahertz band (0.1-10 THz) communications is one of the candidates for 6G systems due to intrinsic massive bandwidth and data rate support. Having demonstrated the significant potential of THz band at various atmospheric altitudes, in this article, we discuss the prospects of THz communications for drone networks, more specifically, for Drone Sensor Networks (DSNs). For 6G non-terrestrial communication scenarios, drones will not only serve as on-demand base-stations, as supporting alternatives or backhauls for the terrestrial base stations, but they will also provide seamless connectivity for distributed monitoring and surveillance applications, which require an ultra-reliable low latency service for carrying multimedia data. THz band sensing will also provide additional sensing capabilities from the sky to THz-enabled DSNs. Presenting this vision, in this paper, we first discuss possible use cases of THz-enabled drone networks considering communication, sensing and localization aspects. Then, for revealing the capacity potential of THz-enabled drone networks, we provide motivating channel capacity results for communication of drones at different altitudes, under ideal channel conditions with no fading and realistic channel with beam misalignment and multipath fading. We further present major challenges pertaining to employing the THz band for DSNs, addressing physical layer issues, followed with spectrum and interference management, medium access control and higher layers and security, while reviewing some prominent solutions. Finally, we highlight future research directions with Artificial Intelligence (AI)/Machine Learning (ML)-based approaches and mobile edge computing.*

Keywords – Artificial intelligence, disaster management, drone networks, drone sensor networks, machine learning, mobile edge computing, monitoring, surveillance, terahertz communications, terahertz sensing

1. INTRODUCTION

Drones will soon inhabit our skies as they are easily available, reliable and low-cost devices. The demand for such hovering drones is increasingly witnessed in civil and government applications, as globally, many governments and industries have been investing heavily in deploying drone networks as per their requirements [1]. Typically, small drones with multi-copter-like functionalities are favo-rable due to their cheap maintenance and convenient deployments [2]. In order to achieve a certain mission, it is usually desirable to deploy a collection or swarm of drones in a networked fashion [3]. Such drone networks or Drone Sensor Networks (DSNs) can monitor a large coverage area and the sensed data can be gathered with enhanced reliability, resilience and fault tolerance under diverse conditions.

DSNs can be highly viable in many real-world scenarios: For military surveillance applications (Fig. 1(a)), DSNs can monitor a sensitive area, such as across international borders, where highly delicate military data (in the form of images or videos) can be transmitted securely. In addition to the communications perspective, DSNs involving drones with sensing and processing capabilities can be utilized in some applications, such as disaster management, for instance for detecting dangerous gases (Fig. 1(b)). Moreover, drone networks with drones with base station-like capabilities can provide seamless on-demand network coverage from the air (Fig. 1(c)), as an alternative as well as support to the terrestrial base stations, for communal gatherings, concerts etc. Such drone base stations can also be useful in disaster-struck areas, where the terrestrial communication infrastructure is damaged. Nevertheless, in each of the above-mentioned possible real-world applications, frequent mobility of the drones with constrained energy resources will be required to be addressed to achieve the desired/optimal performance.

The intelligent information society of 2030 is expected to be globally information driven, highly digitized, with the support of unlimited and near instant complete wireless connectivity [4]. 6G, here, will be the prime catalyst for achieving this target, connecting everything, including wireless coverage in all dimensions, as well as concatenating almost all different functions such as, communication, sensing, imaging, computing, caching, navigation (e.g., radars), control, for supporting nearly all real-world applications [5]. As wireless communications are rapidly progressing towards 6G, from the exponentially growing network traffic arises need of exploiting the electromagnetic spectrum above the existing sub 6 GHz bands, which are almost saturated. A possible solution to this need is to utilize the Terahertz (THz) band (0.1-10 THz) [6], as the bridge between the 5G millimeter wave band and the free space optics band [7, 8]. A THz band offers huge bandwidth, favorable for very high data rate

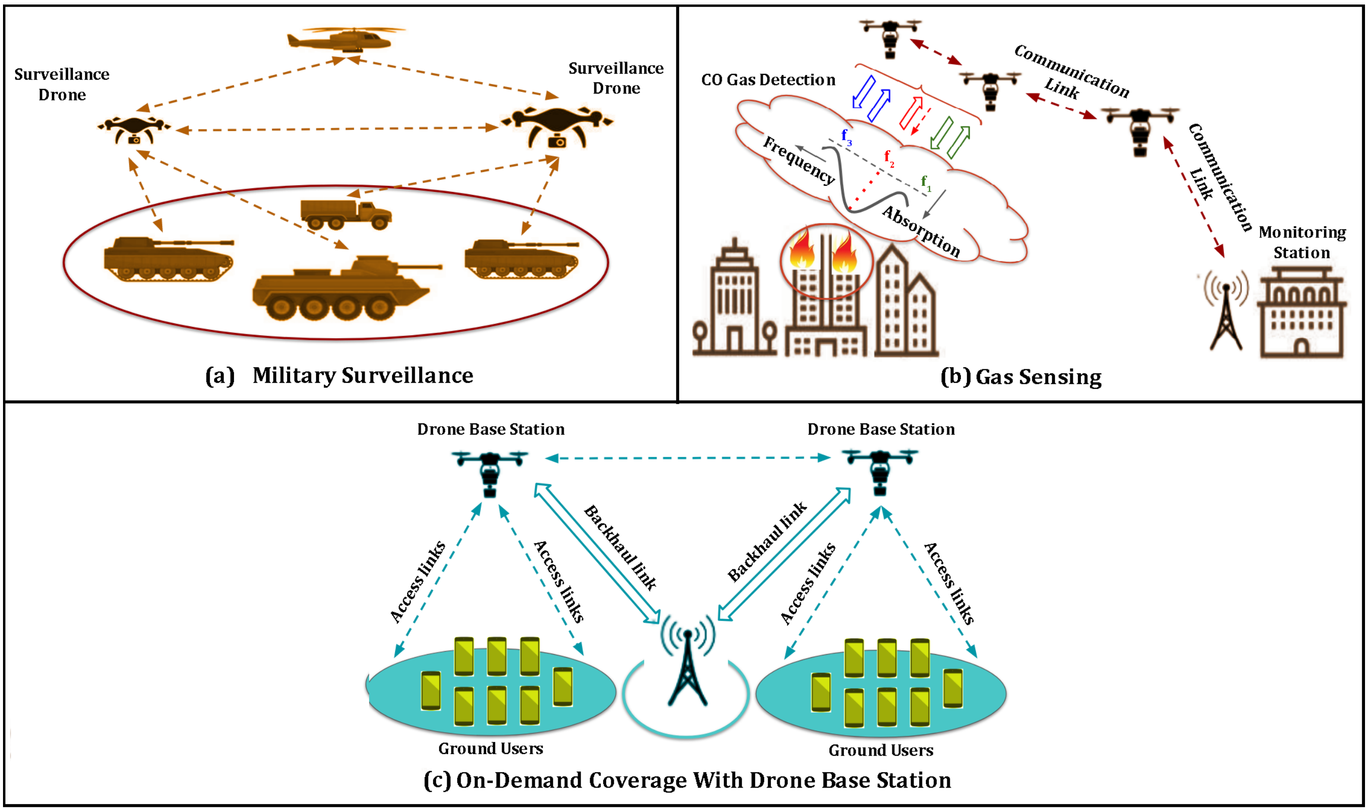


Fig. 1 – Possible real-world scenarios of THz-enabled drone networks.

applications, while at the same time promises massive antenna gains due to shorter wavelengths [9, 10]. However, THz communications are restricted by the absorption loss, which is highest at the sea-level, as the atmospheric gas concentration is at the maximum [11, 12]. Consequently, THz communications have been studied mainly for short transmission distances at the sea-level, such as for on- chip communications [13] or for connecting data centers within up to 10 m [14].

THz bands have been recently considered for aerial communications. Despite the highly mobile nature of aerial vehicles, a THz band, due to very high frequency, promises minimized Doppler effect, making massive rate communication links realizable within mobile aerial vehicles by optimal selection of the beam patterns [15]. For the razor-sharp beams due to carrier frequencies in the order of THz, the communication links between hovering aerial vehicles have to be highly aligned. The influence of the micro, small and large-scale mobility uncertainties for drones communicating in millimeter wave and THz bands are studied in [16, 17]. It is shown that without adaptive beam-width control, micro-scale mobility induces negligible link capacity degradation, whereas small-scale mobility and large-scale mobility can induce significant degradation in the link capacity, with larger outages. THz-based drone (Unmanned Aerial Vehicle, UAV) networks are analyzed in [18] by assessing coverage probability and area spectrum efficiency. It is concluded that due to massive path loss incurred by THz band waves at 0.35 THz, a larger UAV density is required for a certain coverage probability, as compared to lower carrier frequencies. In [19], THz MIMO-OFDM communication between two UAVs is studied by analyzing the orientation and position estimation error bounds. It is shown that the positioning accuracy at the millimeter levels can be achieved provided that the transmitter to receiver separation is considerably small.

In [11], we have performed path loss and total usable bandwidth analyses over a THz band (0.75-10 THz), considering constant narrowband noise approach for four types of aerial vehicles at different altitudes: Drones (at 1 km), jet planes (at 10 km), high altitude UAVs (at 16 km), and satellites (at 99 km). The path loss analysis shows that the absorption effect diminishes at the higher altitudes and the total path loss behaves as free space spread loss. Moreover, the total usable bandwidth analysis infers that at the higher altitudes (e.g. high altitude UAVs and satellites), the entire THz band (0.75-10 THz) becomes feasible as a single transmission window, which is about 9.25 THz wide. In our subsequent work [12], we have proposed an alternative channel model for THz communications, where, by taking the colored nature of noise Power Spectral Density (PSD) into account, the commonly flat bands in path loss (gain) and noise PSD are determined for the THz spectrum (0.75-10 THz), and a variable bandwidth capacity computation method is proposed as an alternative to the standard capacity computation. Our extensive capacity analysis of the same four aerial scenarios has shown that under fading channel conditions, the ergodic capacity at the sea-level is enhanced by an order of magnitude for the drones hovering at 100 m altitude supporting multiple Tbps at a range of 10 m, whereas 10s of Tbps are realizable among jet planes (10 km altitude) and the high altitude UAVs (16 km altitude), and multitude of 100s of Tbps are also achievable for inter-satellite links (also cubesats) at a range of 1 km. In both works [11, 12], it is concluded that the THz communications is highly viable for non-terrestrial communications. Motivated by our earlier analyses for THz communications among aerial vehicles in [11, 12], in this survey, we review THz band communications and sensing for drone networks, more specifically for DSNs and we highlight the prospects. First, we present applications of THz-enabled DSNs. Secondly, we present capacity results for drone-to- drone links, considering drones at different altitudes and varying fading conditions. We show that THz communi- cations can be quite promising even under realistic fading conditions, providing up to many Tbps at 10 m range and 10s of Gbps at 50 m range. Motivated by the capacity potential, we next discuss the open issues for THz communications in DSNs, along with research directions. Lastly, we highlight Artificial Intelligence (AI) and Machine Learning (ML)-based approaches.

2. APPLICATIONS OF THZ BAND DRONE NETWORKS

With the THz band being one of the key enablers of prospective 6G research, we envision employing the THz band for drone networks as expected use cases of 6G non-terrestrial networks. Fig. 2 depicts such possible applications of THz-enabled drone networks, namely monitoring and surveillance, sensing, localization and on- demand network coverage, as elaborated next. It is worth-mentioning here that within each of the following applications of the THz band communications, sensing and localization, it will be important to deploy the THz- enabled drones strategically [6] so as to fully exploit the THz band for drone networks and DSNs. These strate- gies include Ultra-Massive Multiple Input Multiple Out- put (UM-MIMO) and Reconfigurable Intelligent Surfaces (RIS), which are explained in Section 4 of this paper, in detail.

2.1 Monitoring/Surveillance

As 6G research is progressing towards ubiquitous automation, aerial monitoring using drones has become increasingly popular [20]. Networked state-of-the-art drones will become an essential aerial resource for many real-world monitoring applications, such as military surveillance, disaster management, etc. In the following, we cover such practical monitoring applications for THz-enabled drone networks.

Military Applications

Drone networks have become an integral part of military applications around the globe. With the aid of quickly deployable drone networks, a multitude of military-related activities can be performed effectively. Such activities include border patrolling by monitoring high resolution real-time videos. For instance in [21], the concept of *BorderSense* is introduced, which is a new hybrid concept of wireless sensor networks, including underground sensors, on-ground sensors, and drones for wireless sensing from the air. With the *BorderSense* framework, drones can provide mobility freedom together with on-board high resolution cameras and highly sensitive sensors for provisioning an enhanced coverage as per the military requirements. Here, for the military applications, bandwidth should be provisioned substantially together with an ultra-reliable low latency service in order to transfer sensitive military data/information within fractions of a second [10]. In addition, drones providing such military applications (e.g., air to air and air to ground) will need to be highly energy efficient as it will not be realistically possible to frequently replace their on-board power sources (batteries). With THz-enabled DSNs, a massive THz bandwidth will promise high resolution video monitoring of sensitive areas, as in *BorderSense* framework. Moreover, thanks to the unique propagation characteristics of the THz band, THz waves show a unique reflective nature to the metallic surfaces. Therefore, a THz band can be effectively utilized in the military surveillance applications such as detecting certain weapons [22]. The detection of any possible explosives is also possible using THz spectral imaging [23]. For instance, THz spectral imaging can detect small land mines, while the ground radars cannot distinguish between such mines and rocks [24]. THz-enabled DSNs with mobility freedom can assist a military by detecting such mines without any direct human contact.

Disaster Management

Drone networks or DSNs can be considered in various disaster management applications, e.g., for Early Warning Systems (EWSs), which perform environmental and structural monitoring and process the collected informa- tion for disaster predictions. Such a DSN can also support Search And Rescue (SAR) missions in the post disaster management scenarios. DSNs can be utilized to restore the damaged communication infrastructure in case of a disaster. Additionally, post-disaster damage assessment can be made possible using DSNs with video monitoring [25]. THz-enabled drone networks and DSNs with massive bandwidth can support an aerial communication backup for EWS, SAR over the disaster-struck region. Energy efficiency of the drones will be a critical parameter for such autonomous aerial operations [26], where frequent replacement of the on-board batteries may not be a feasible operation.

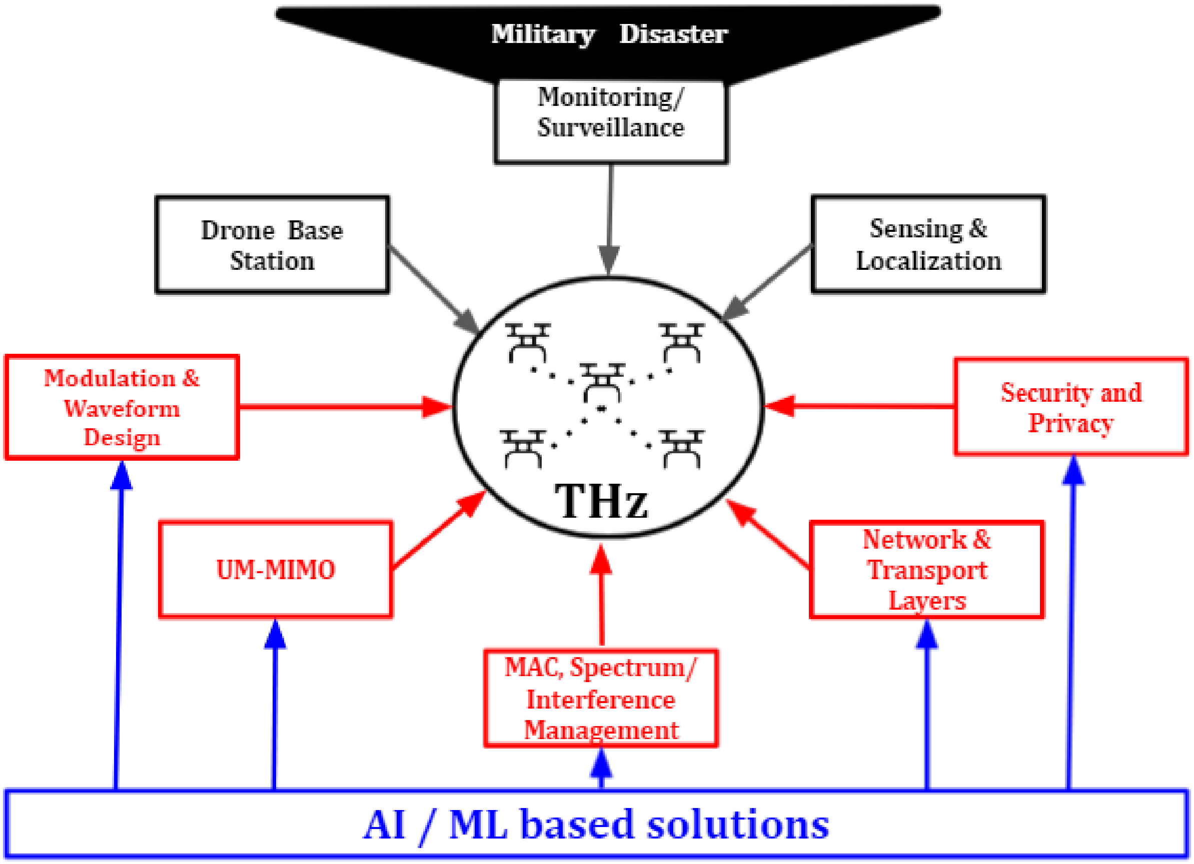


Fig. 2 – THz-enabled drone sensor networks: Applications, challenges, and solutions.

2.2 Terahertz for sensing

Unlike infrared and microwaves, THz waves hold many unique transmission properties such as a quasi-optical nature and molecular absorption, making the THz band an appropriate candidate for some sensing applications, such as gas and material sensing, quality control, chemical and biological sensing, bearing unique THz spectral signatures. Hence, THz-enabled DSNs can be deployed in a variety of practical sensing applications. THz signals possess distinctive spectral signatures, i.e., quasi-optical nature having varying absorption loss across the THz band, due to the molecular rotations/ vibrations, promising that THz waves can be highly suitable for applications related to the rotational spectroscopy, such as gas detection [6, 22]. Consequently, it becomes possible to utilize THz waves for the purposes of detecting various poisonous/dangerous gases like hydrogen cyanide, carbon monoxide, etc., possibly over disaster-struck places, including fire eruption, massive earthquakes, volcanic eruptions, etc. THz-enabled DSNs, here, can be employed to promptly detect/sense such poisonous gases with the aid of THz waves by rapid detection and relaying such vital information, saving any further victims and avoiding other secondary disasters. Both active and passive sensing methods can be employed for poisonous gas detection using THz spectroscopy. An example of detecting a dangerous gas, such as carbon monoxide is illustrated in Fig. 1. Here, a THz wave with spectrum ranging from f_1 to f_3 is transmitted through a drone Tx, which is then received by the drone Rx. If the dangerous gas pertaining an absorption at f_2 is available, the received THz wave level at f_2 shall be lower as compared to the other frequencies. Thus, the gaseous presence is sensed/detected. For the gas sensing, THz band of 0.2-1 THz has been considered in the literature [27]. After detecting the gas successfully, again, the THz band can be utilized for relaying (communicating) the sensed information via drone networks to a nearby monitoring station for rapid rescue operations. Thus, an all-in-one operation (sensing and communication) can be conveniently achieved by deploying THz-enabled DSNs and drone networks.

2.3 Localization

Besides other requirements of the prospective 6G systems, high precision localization is also expected. It is predicted that the THz band will promise localization accuracy in the order of centimeters [28], which is far better than the existing localization-based ser- vices, such as Global Positioning System (GPS), and cell- multilateration. Localization methods at higher frequencies, such as across the THz band rely on the technique of Simultaneous Localization and Mapping

(SLAM), where the overall accuracy is enhanced by obtaining very high resolution images of the surroundings (environment). SLAM techniques comprise of three main stages: 1) image capturing of the surroundings, 2) range estima- tion from the user, and 3) aggregation of the images at the approximated ranges. For example, an accuracy of the sub-centimeter levels can be acquired by making 3- dimensional images of the surroundings with the aid of THz signals and projecting the time and angle of arrival details from the user for estimating the locations. Instead of ground-based localization, THz-enabled DSNs can assist localization applications with an improved aerial field of view over a larger coverage area [22].

2.4 Drone base stations

The concept of Drone Base Stations (DBS) has recently emerged in the literature, e.g., [29, 30, 31, 32], where a hovering/mobile drone with base station-like capabilities can serve the ground users as an alternative back-up to the terrestrial network. DBS has already been investigated in [33] as an extension of 5G. However, there is some recent work on DBS for 6G networks and tech- nologies. For instance in [7], DBSs have been considered as a subset of Wireless Infrastructure Devices (WIDs), promising coverage and capacity improvements for the prospective 6G networks. In [34], DBSs have been recognized as a major challenge related to the intelligent handover of drones in multiple DBS networks for 6G technologies. By deploying DBS, aerial wireless coverage can be provided to the ground users as well as to the hot spots, especially in the areas/regions with scarce or no communication infrastructure [35]. Also, it will be possible to provide seamless high rate connectivity for real-time multimedia streaming and/or for on-demand applications, such as in concerts [36, 37], reducing the overall communication load of a nearby terrestrial base station. Moreover, in case of a natural disaster, multiple DBSs can be deployed over the entire disaster-struck area with the damaged communication infrastructure, e.g., macro-hot spots [29], thereby providing high rate communication backup from the air. Here, for DBS scenarios, bandwidth should be adequately provisioned with an ultra-low latency service to provide uninterrupted coverage to the ground users. Thanks to a large THz bandwidth, THz-enabled DBSs can support each of the above use cases effectively. Still, THz-enabled DBSs will also take into account frequent handovers between adjacent DBSs for providing seamless coverage to the ground users [34].

3. CAPACITY OF THZ COMMUNICATION AMONG DRONES

Having discussed the insights into the possible applications of the THz-enabled drone networks, in this section, we present THz channel capacity analysis and some results for practical drone scenarios as our motivation for considering THz communications in DSNs.

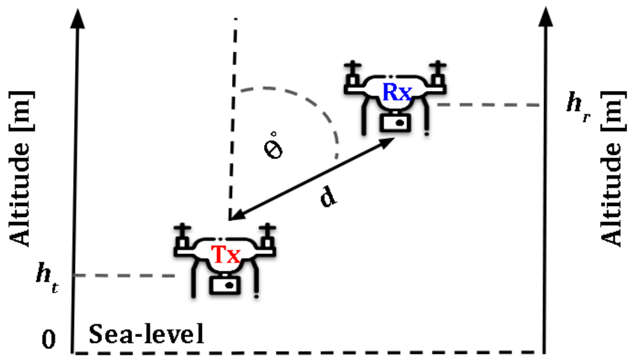


Fig. 3 – System model of a THz-enabled drone network at various atmospheric altitudes consisting of a transmitter drone at altitude h_t, and a receiver drone at altitude, h_r. For a given h_t, h_r is found according to θ and d.

Fig. 3 illustrates a simple system model, where a transmit- ter drone, Tx, and a receiver drone, Rx are hovering at alti- tudes h_t and h_r, respectively. θ is the phase from the verti- cal axis, determining the direction of communication. For instance, $\theta = 0^\circ$ corresponds to the vertically-up communi- cation, 90° denotes horizontal direction, while 180° is the vertically-down direction.

As discussed in Section 1, THz waves experience a dominant water vapor absorption gain in addition to the free space spread gain. Besides, the Beam Misalignment (BM) fading and Multipath (MP) fading also contribute to the overall gain obtained as follows [12, 38]:

$$G = G_l G_p G_f, \quad (1)$$

where, G_l is the path gain coefficient, and G_p and G_f refer to BM and MP fading, respectively. G_l includes the free space spread gain coefficient, G_s, and the absorption gain coefficient, G_a given as:

$$G_l(f, h_t, h_r, d) = G_l(f, h_t, \theta, d) = G_s(f, d) G_a(f, h_t, \theta, d)\ , \quad (2)$$

where f denotes the carrier frequency of the THz wave in hertz, h_t, h_r are the transmitter (Tx) and receiver (Rx) altitudes in meters, respectively. d is the transmission distance between Tx and Rx, and θ denotes the relative position of Rx with respect to Tx as depicted in Fig. 3. $G_s(f, d)$ is the free space spread gain coefficient due to the THz wave attenuation as it propagates across the atmosphere via an isotropic antenna, obtained as:

$$G_s(f, d) = \frac{c}{4\pi f d}\ , \quad (3)$$

where c is the speed of the THz wave in free space, i.e., 299,792,458 m/s.

$G_a(f, h_t, \theta, d)$, the absorption gain coefficient is mainly induced by the water vapor molecules present in the atmosphere, mathematically expressed as,

$$G_a(f, h_t, \theta, d) = \left(\tau(f, T(h_t, \theta, d), v(h_t, \theta, d))\right)^{1/2}\ , \quad (4)$$

where $\tau(f,\, T\,(h_t,\, \theta,\, d),\, v(h_t,\, \theta,\, d))$ is the atmospheric medium's transmittance. We employ Line-By-Line Radiative Transfer Model (LBLRTM) for obtaining realistic transmittance values across various atmospheric altitudes [39, 40]. $T\,(h_t,\, \theta,\, d)$ and $v(h_t,\, \theta,\, d)$ are the atmospheric temperature (in Kelvin) and water vapor concentration (in %), respectively, for the link between the Tx drone hovering at h_t and Rx drone at h_r.

In regard to the impact of the mobility of the communicating drones and the resulting Doppler spread, thanks to the very high operating frequency in the order of THz, mobile drones observe minimized Doppler effect [15], promising high rate links between communicating drones. It has been shown in [41, 42] that for a typical drone relative velocity, v_r = 10 m/s, the maximum Doppler shift is negligible. As an example, by considering f_c = 0.75 THz, and v_r = 10 m/s, the maximum Doppler shift is: $f_{d.max}$ = $f_c.v/c$ = 25017.31 Hz, which is negligible in terms of the inter-carrier interference. Moreover, since we are con- sidering drone to drone communications, where a swarm (group) of drones move together, v_r can be less than 10 m/s (up to 0 m/s) and $f_{d.max}$ will be even smaller than the example provided above. Therefore, we neglect the effect of the Doppler spread in our capacity computations.

For beam misalignment fading, we consider the following probability density function for the BM fading coefficient, G_p, [43]:

$$f_{G_p}(x) = \frac{\zeta^2}{{A_0}^{\zeta^2}} x^{\zeta^2-1}\,, \tag{5}$$

where $\zeta = \frac{w_{eq}}{2\sigma_s}$, w_{eq} is the equivalent Tx beam width, σ_s denotes the jitter (BM) standard deviation, and A_0 is the fraction of power collected at Rx at no beam misalignment. This BM fading model has been widely employed in many studies of free space optical systems. For more details of the BM fading model, we refer to our earlier work [12] and also [43].

Finally, for incorporating multipath fading, we consider the famous α-μ model as follows [44]:

$$f_{G_f}(x) = \frac{\alpha\mu^{\mu}}{\hat{G}_f^{\alpha\mu}\Gamma(\mu)} x^{\alpha\mu-1} exp\left(-\mu\frac{x^{\alpha}}{\hat{G}_f^{\alpha}}\right)\,, \tag{6}$$

where $f_{G_f}(x)$ is the pdf of the MP fading coefficient, G_f, α is the fading parameter, μ is the normalized variance of the channel envelope under fading, and G_f is the α- root mean value. The α- μ model is a common model of several famous fading distributions. For instance, $\alpha = 2$ and $\mu = 1$ represents Rayleigh fading, etc.

For computing noise power, P_n, we consider a constant narrowband approach [14] across the THz band (0.75-10 THz), where each narrowband, Δf is 0.3 GHz wide, which is the spectral resolution of LBLRTM. Numerically,

$$P_n(f, h_t, \theta, d, \Delta f) = k_B \int_{\Delta f} T_{noise}(f, h_t, \theta, d) df, \tag{7}$$

where k_B is the Boltzmann's constant, T_{noise} is the molecular noise temperature obtained as $T_{noise}(f,\, h_t,\, \theta,\, d) = T_0\epsilon(f,\, h_t,\, \theta,\, d)$. Here, T_0 is the reference temperature (in Kelvin), and ϵ is the channel's emissivity, $\epsilon(f,\, h_t, \theta, d) = 1 - \tau(f,\, T\,(h_t,\, \theta,\, d),\, v(h_t,\, \theta,\, d))$ [14]. Hence, T_{noise} is a function of transmittance, τ, which is obtained using LBLRTM. In the results provided in this paper, for capturing the absorption effect across the THz band (0.75-10 THz), US Standard 1976 weather profile is set in LBLRTM [12]. For computing the capacity of THz drone-to-drone links under ideal, no fading channel, the total channel gain is set as, $G = G_l$. In this paper, we consider the standard narrowband capacity computation in [14, 45]:

$$C(h_t, \theta, d) = \sum_{k=1}^{K} \Delta f\, log_2 \left[1 + \frac{P_T^k\, |G(f_k, h_t, \theta, d)|^2\, G_T}{P_n(f_k, h_t, \theta, d, (\Delta f))}\right]\,, \tag{8}$$

where total transmit power, P_T, and total antenna gain (from Tx and Rx antennas), G_T are set to practical values as P_T = 24 dBm (0.25 W) [46] and G_T = 60 dBi [47, 48], respectively. For power allocation, we consider both the Water-Filling (WF) and Equal-Power (EP) schemes [14]. In WF allocation, the total transmit power, P_T is optimally distributed across the THz band (0.75-10 THz) comprising of constant narrowbands, $k = 1,\, 2,\, 3,\, \ldots,\, K$, each 0.3 GHz wide (i.e., LBLRTM's spectral resolution), as:

$$\frac{P_T^k}{P_T} = \begin{cases} \frac{1}{\gamma_\circ} - \frac{1}{\gamma_k} & , \gamma_k \geq \gamma_\circ \ \text{ s.t. } \sum_{k=1}^{K} P_T^k \leq P_T \\ 0 & , \gamma_k < \gamma_\circ\,, \end{cases} \tag{9}$$

where P_T^k is the optimal power for the constant narrowband, k, $\gamma_\circ$ is the threshold SNR, γ_k is the SNR of k. $\gamma_\circ$ is obtained by $\sum_{k=1}^{K}\left(\frac{1}{\gamma_\circ} - \frac{1}{\gamma_k}\right) = 1$ [49].

In EP allocation, P_T is distributed equally within all k across the entire THz band (0.75-10 THz) [12].

For the channel under fading, involving BM and MP fading, the channel gain is set as $G = G_l G_p G_f$, and we evaluate the ergodic capacity by averaging results over 100 realizations, as follows:

$$\overline{C(h_t, \theta, d)} = \Delta f\, \mathbb{E}\left(\sum_{k=1}^{K} log_2 \left[1 + \frac{P_T^k\, |G(f_k, h_t, \theta, d)|^2\, G_T}{P_n(f_k, h_t, \theta, d, (\Delta f))}\right]\right), \tag{10}$$

where $\mathbb{E}(.\ )$ denotes the expectation taking over channel realizations under fading.

Next, we present the capacity and ergodic capacity results specifically for drone scenarios, considering various practical settings of Tx and Rx drone altitudes, zenith angles and transmission ranges. Fig. 4(a)-(c) depict the channel capacity as the function of transmission range under ideal, i.e., under no fading channel. The capacity results with no fading (ideal) channel are included in our analysis as the benchmark to compare how much of the capacity is degraded when realistic beam misalignment fading and multipath fading are introduced into the channel, as provided in the subsequent discussion. Three drone Tx altitudes, h_t, are considered i.e., h_t = 100 m, 500 m, and 1 km, whereas h_r i.e., the Rx drone altitude for each setting is obtained using h_t, θ (angle in degrees between Tx and Rx drones), and d.

It can be seen for a given h_t, changing θ, i.e., the direction of communication from 0° (vertically-up) (Fig. 4(a)) to 90° (horizontal) (Fig. 4(b)) down to 180° (vertically-down)(Fig. 4(c)) do not incur considerable variations in the ca- pacity. This is due to the dense and homogeneous atmosphere across lower atmospheric altitudes. Nevertheless, increasing h_t shows promising capacity improvements. For instance, at h_t = 100 m, 0°, and d = 100 m, capacity values correspond to 505.8 Gbps and 34.52 Gbps with WF and EP allocation schemes, respectively. For the identical θ and d settings but at a higher Tx drone altitude, h_t = 1 km, the capacity values stand at 652.6 Gbps and 62.89 Gbps with WF and EP allocations, respectively. This is because traversing up across the atmosphere from 100 m to 1 km observes substantial decrements in the water vapor concentration levels [12], which can be highly leveraged in drone networks communicating over the THz band. Additionally, for overcoming the distance issue observed across lower atmospheric levels, for instance, for h_t lower than 100 m, multiple drones can be deployed sufficiently close to each other, in a networked fashion, where they can be treated as relays. These results showcase the massive capacity potential of the THz band for drone networks, promising links in the order of up to several 10s of Gbps using EP allocation, and up to many 100s of Gbps with WF power allocation for transmission ranges up to 100 m. Fig. 5 depicts the ergodic capacity trend for short range, i.e., d = 10 m, under BM fading and MP fading parameters [12]. Interestingly, it can be seen in Fig. 5(a) that for MP fading parameter, μ = 1, which corresponds to pure NLOS, Rayleigh fading, increasing the normalized jitter standard deviation, σ_s/a does not cause substantial ergodic capacity degradation. Due to nearby reflections the NLOS MP fading components are more dominating than the BM fading components for short range; hence severeness of BM does not affect the ergodic capacity. Meanwhile, Fig. 5(b) shows that for fixed BM fading parameter, $\sigma_s/a = 5$, MP fading degrades the ergodic capacity by 26 % for EP allocation, and 16.5 % for WF allocation, as the MP effect is varied from μ = 10, indicating a strong LOS along with NLOS components to μ =1, i.e., pure NLOS Rayleigh fading. Next, we present the achievable ergodic capacity at d = 50 m, under variable BM fading with Rayleigh MP fading (μ = 1) in Fig. 6(a), and variable MP fading with σ_s/a = 5 Fig. 6(b). For this range, increasing σ_s/a from 1 to 10 decreases the ergodic capacity substantially, e.g., by an order of magnitude for EP allocation, as σ_s/a is increased from 1 to 10. On the other hand, decreasing μ, from 10 to 1 with given σ_s/a = 5 shows no considerable change in ergodic capacity. The ergodic capacity results in Fig. 5 and Fig. 6 depicting that at short ranges, it is the MP fading that mainly affects ergodic capacity, while at long range, it is mainly the BM fading. With this analysis, we emphasize that THz band communication in drone networks can promise massive rate links even under realistic BM fading and MP fading con- ditions. We refer the readers to [12] for an in-depth ca- pacity analysis of THz communications for drones and the other three aerial vehicles, where both standard narrowband and variable bandwidth capacity computa- tions are considered for various altitudes, distances, posi- tions/orientations of the vehicles (i.e., the entire range of θ from 0° to 180°) by leveraging LBLRTM for THz absorp- tion gains, evaluating no fading, BM fading and MP fading conditions.

4. OPEN ISSUES AND RESEARCH DIRECTIONS

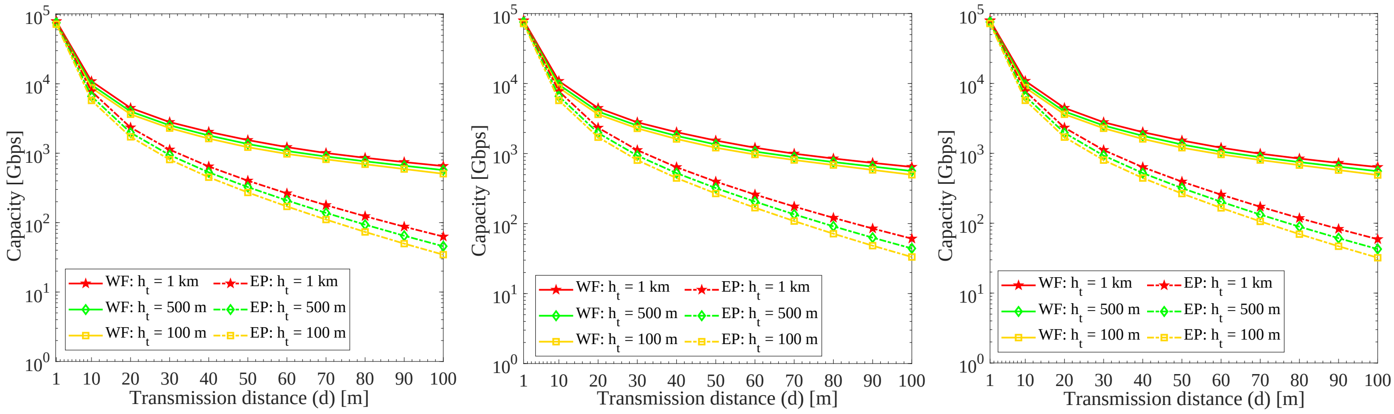
Design and implementation of THz-enabled drone networks and DSNs require novel communication schemes and networking protocols, including but not limited to modulation and waveform design, ultra-massive Multiple Input Multiple Output (MIMO), spectrum and interference management, Medium Access Control (MAC) and higher network layers, security and privacy issues.

4.1 Physical layer

THz band drone communications primarily requires enhanced THz band channel models. For this purpose, measurement-based studies need to be pursued at drone altitudes in various propagation environments and under drone mobility scenarios, so that the existing line-of-sight and non line-of-sight models with beam misalignment and generic multipath fading (as considered in this work) can be improved with specific stochastic channel models for THz links among drones. A recent work on active and passive THz systems is presented in [50], where measurement results at 140 GHz (0.14 THz) have been provided for rooftop surrogate satellite systems and terrestrial networks. Based on the enhanced channel models, modulation and waveform design should be tailored for THz band communications in drone networks or DSNs.

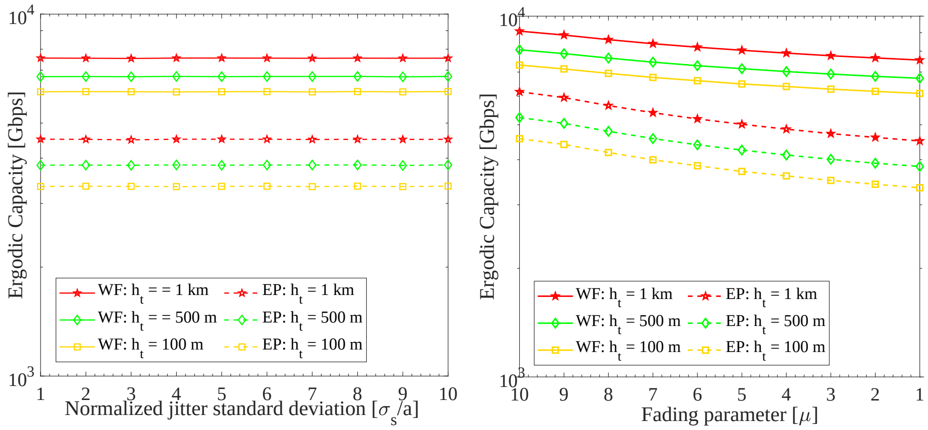
Modulation

The state-of-the-art modulation schemes that can be potentially employed for THz band communications include Single-Carrier (SC) modulation, multi-carrier modulation, Orthogonal Frequency Division Multiplexing (OFDM), Cyclic Prefix Orthogonal Frequency Division Multiplexing (CP-OFDM) and even Non-Orthogonal Multiple Access (NOMA). In what follows, we discuss each of the aforementioned modulation schemes in the perspective of THz band communications for drone networks. Non-overlapping transmission windows are termed as Single Carrier (SC) modulation, having some provision of the carrier aggregation [51]. However, due to the intrinsic frequency-selective nature of THz channel, multi-carrier modulations would also help in some form of carrier aggregation with multiple individual/ non-overlapping single carriers [52]. The implementation of practical THz transceivers is another challenging task,



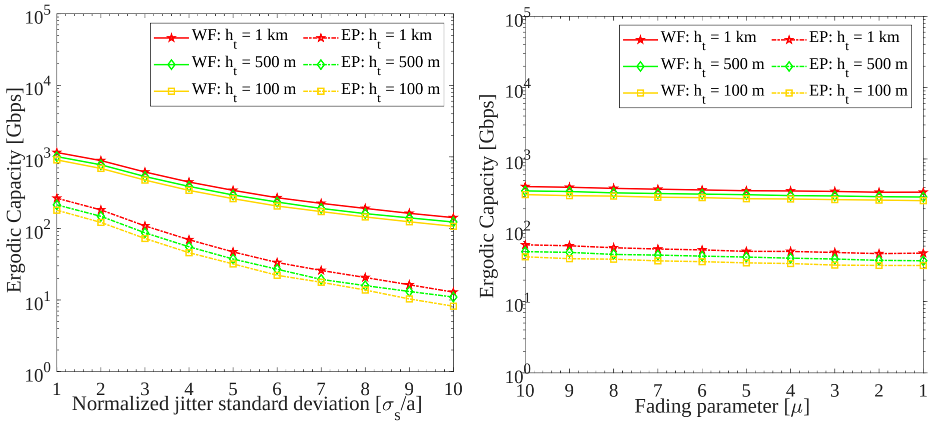
(a) Vertically-up communication, θ = 0°. **(b)** Horizontal communication, θ = 90°. **(c)** Vertically-down communication, θ = 180°.

Fig. 4 – Capacity under no fading as a function of distance, for various drone transmitter altitudes and directions of communication (vertically up, horizontal and vertically down, i.e., $\theta = 0°$, $90°$, and $180°$, respectively). For a given h_t and d, varying direction (θ) does not considerably affect the capacity, as the atmosphere is dense and homogeneous across for drone altitudes up to h_t = 1 km.



(a) MP fading parameter, μ = 1. **(b)** Normalized jitter standard deviation, σ_s/a = 5.

Fig. 5 – Effect of beam misalignment fading and multipath fading on the ergodic capacity: θ = 90° (horizontal communication), and d = 10 m across various drone transmitter altitudes. For this range, MP fading shows more dominant impact on the ergodic capacity as compared to BM fading due to short transmission distance.



(a) MP fading parameter, μ = 1. **(b)** Normalized jitter standard deviation, σ_s/a = 5.

Fig. 6 – Effect of beam misalignment fading and multipath fading on the ergodic capacity: θ = 90° (horizontal communication), and d = 50 m across various drone transmitter altitudes. For this case, BM fading shows more dominant impact on the ergodic capacity as compared to the MP fading, due to a large transmission distance.

as the conventional RF circuitry cannot support data rates in the order of several 100s of Gbps or Tbps as provisioned in Section 3. Moreover, novel signal processing techniques will be required to counter the mismatch between the state-of-the-art digital baseband systems and the large bandwidth offered by the THz band [52]. Recently, there have been advancements related to the THz transceivers; both in the electronic and the photonic domains [22]. The advancements towards practical THz transceivers (both electronic and photonic) have been well summarized in [53]. For short-range communication (below one meter), impulse-radio-like communication based on one-hundred-femtosecond-long pulses following an on-off keying modulation spread in time has been proposed in [36]. Such very short pulses, which are already utilized as the basis of many THz sensing systems, can be generated and detected with current technologies. For longer communication distances, new dynamic bandwidth modulations are required for not only overcoming but also leveraging the unique distance-dependent bandwidth created by molecular absorption [11, 12]. Orthogonal Frequency Division Multiplexing (OFDM) has widely been implemented in broadband wireless systems since 4G for achieving higher spectral efficiency. In [54], OFDM is proposed for 60 GHz millimeter wave systems. For 5G, several wireless standards including Long Term Evolution (LTE), Wireless Fidelity (Wi-Fi), Asynchronous Digital Subscriber Line (ADSL) etc., have adopted Cyclic Prefix Orthogonal Frequency Division Multiplexing (CP-OFDM)[55]. Multiple communicating nodes, each with Tbps of data (which would also be the case in THz-enabled DSNs) would require relaying data in an asynchronous manner. Also, in such asynchronous multiple user access, the subcarriers with CP-OFDM do not remain orthogonal, which introduces substantial inter-carrier interference [56]. This makes CP-OFDM infeasible for DSNs. Nevertheless, OFDM systems promise utilization of the non-overlapping spectrum for the improved spectrum efficiency as compared to the pulse-based communication systems [57, 58]. However, implementation of OFDM transceivers in the THz band is especially complex due to stringent frequency synchronization requirements, with the sampling rates in the typical order of several Giga samples/sec or even Tera samples per second. Additionally, large high Peak-to-Average Power Ratio (PAPR) also makes OFDM implementation not feasible over the THz band [52, 59]. In recent years, Non-Orthogonal Multiple Access (NOMA) has gained considerable attention, as it promises not only greater link rates for both the downlink and uplink transmissions, but it also provisions a way to counter the packet collision issue, e.g., in MTC with grant-free access [60]. NOMA is adopted for THz systems in [61] by making use of the frequency and distance-dependent THz spectrum. The concept of hybrid beamforming is proposed for forming user clusters, and NOMA-based grouping and Long-User-Central-Window (LUCW)-based sub-band allotment within a user cluster are proposed for improving user fairness as well as spectral efficiency. For mitigating the water vapor-based absorption effect in the THz band, the conventional modulation schemes can be further optimized. For this purpose, in [51], distance-aware multi-carrier schemes are proposed. Resource optimizations include power allocation as in [57], where long range networks are established using a pulse-based multi-wide band waveform design by adapting power allocations over variable number of frames. By adapting the symbol time and modulation order, in [59], a hierarchical modulation scheme is proposed for a system with a single transmitter and multiple receivers, supporting various streams of data for multiple users at variable ranges. In [62], distance-adaptive and bandwidth-dependent modulations using OFDM in THz band are proposed. It is worth mentioning here that, the aforementioned schemes have been proposed for THz band communications at sea level. These schemes will need to be adapted for specifically the drone scenarios or DSNs, considering altitudes as well as drone mobility.

Ultra-Massive MIMO

A main issue for the THz band communications is constituted by the frequency-selective and an extremely high path loss, which simply crosses 100 dB for ranges greater than only a few meters under LOS channels. This path loss is even worse under NLOS channel conditions. Consequently, huge gains by highly directional antennas are required for communicating over ranges greater than a few meters. In this regard, the idea of Ultra-Massive (UM)-MIMO has been proposed [63], where extremely-dense arrays of plasmonic nano-antennas are employed. In lieu of deploying the traditional metallic antennas, meta-materials and nano-materials can be exploited for manufacturing plasmonic nano-antennas, which are sufficiently less than the wavelength of the operating carrier frequency. This unique property of the plasmonic nano-antennas enables them to be packed in massively densed arrays. For instance, for an array with a footprint of 1 mm x 1 mm, a sum of 1024 plasmonic nano-antennas designed for 1 THz carrier frequency can be integrated together, keeping the inter-element distance (spacing) of $1/2$ of the plasmonic wavelength. Similar arrays of the plasmonic nano-antennas can be employed at both the Tx and Rx sides simultaneously for countering the massive path loss issue by: 1) Overcoming the spread loss, by targeting the signal transmission in space, and 2) focusing the bandwidths for the signal transmission within the windows having the least absorption levels. By intelligently inputting the array elements, variations of the modes of operation can be utilized in an adaptive fashion. For instance, in UM-Beam-forming (UM-BF), all of the antennas utilize the identical transmit signal, similar to the case of conventional beam-forming. Such a mode can substantially counter the massive path loss/attenuation at the THz band carrier frequencies, thus supporting communication to larger ranges. Additionally, beam-forming also mitigates the problem of the co-channel interference, also utilizing the freedom of angle diversity by moving the razor-sharp THz beam to specific or targeted directions. Using UM-MIMO, a certain group of arrays/antenna elements can be designated for communicating to a specific user. This special mode utilizes various data streams onto a single carrier frequency, thereby increasing the per user capacity, which is also beneficial when the communication links are operating in a limited bandwidth and a high SNR scheme. This special mode enhances the rate by the virtue of Spatial Multiplexing (SM), provided that the channel matrix of the UM-MIMO has sufficient rank and diversity. Conclusively, any amalgamation across UM-SM and UM-BF is realizable. Additionally, for maximizing the utilization of the THz band, promising Tbps links, multiple transmission windows can be employed simul- taneously. For this purpose, multiband UM-MIMO utilizes different carrier frequencies by tuning electrically the frequency response of the plasmonic nano-antennas. One of the major pros of this multiband UM-MIMO technique is that the data can be processed within a substantially smaller bandwidth, hence, lowering the complexity of the system design with an improved flexibility of the spectrum. In this research arena, novel frequency, space and time modulation and coding methods are required to be proposed for such UM-MIMO communication systems. UM-MIMO can also be leveraged for DSNs due to very large beam-forming gains to overcome the huge path loss over the THz band. Moreover, as a byproduct, razor-sharp THz beams would substantially mitigate the interference among communicating drones in the DSNs. Nevertheless, the design of UM-MIMO systems for DSNs and drone networks should also incorporate the effect mobility of the drones to avoid Tx-Rx antenna beam misalignment and maximize the beam-forming. For the case of DBS, implementation of UM-MIMO will be essential to provide aerial coverage to several users at the same time [64]. However, this will be a challenging task, as the flying drones with a limited battery support and single antennas will need to be replaced with UM-MIMO, which will be an important research avenue for the upcoming 6G systems. For instance, recently, THz UM-MIMO communications has been considered for a Space-Air-Ground Integrated Network (SAGIN) comprising of terrestrial, airborne and spaceborne networks [65]. With the plasmonic antenna arrays having nano-antenna spacings, it will be possible to practically implement THz systems onto the flying drones within miniature footprints [52].

Reconfigurable Intelligent Surfaces

Recently, novel tunable metasurfaces are referred to as Reconfigurable Intelligent Surfaces (RIS), which can be used for controlling and optimizing the wireless channel environment [66, 67, 68, 69]. Generally, strong NLOS signals having specular reflections take the surfaces of the existing building infrastructures as electric mirrors, particularly at considerably miniature wavelengths across the THz band. However, RIS built from the metasurfaces and discrete element semiconductors also enables user-customized settings. A similar RIS can sufficiently increase the THz signal power by reflecting the THz signals towards a specific direction. This can be achieved by introducing required phase shifts of the discrete elements in the RIS. In addition, a sufficiently large RIS supporting the aforementioned features can be acquired within miniature footprints at high frequencies, such as the THz band [22, 70, 71]. RISs have already been considered for improving the coverage performance of THz indoor communications at the sea level, as in [72], where the authors have proposed a suboptimal search scheme for the RIS phase shifts. Additionally, RISs have also been consi-dered for THz drone communications [73]. For the drone networks at the low altitude scenarios, e.g., h_t = 100 m and below, up to sea level, THz communication ranges can be substantially extended with the aid of RISs deployed on top of buildings, roofs etc. This can also be achieved for drone-to-ground and ground-to-drone links by placing RISs near the intended user access points [14], Thus, RIS can considerably increase the coverage of a drone network e.g., drone base station, agricultural monitoring in a rural area with a limited or no terrestrial communication infrastructure. etc.

4.2 Spectrum and interference management

With the progression towards 6G, exploitation of the higher frequency bands above existing sub 6 GHz spectrum has become more appealing than ever. This will raise the need of sharing the spectrum using cognitive radio sensing with flexibility [74]. THz spectrum has been identified as a prime communication band for mobile communications within 6G research [28, 75]. Also, 2G to 5G networks across the globe have been utilizing lower frequency bands, which will also be available in 6G networks. Therefore, various spectrum management techniques will be needed for managing the lower, mid and higher frequency bands intelligently. Here, the massive THz band will suffice the spectrum scarcity by assigning different frequency sub-bands to different users subject to different scenarios, mitigating the conventional issue of interference. For instance, in [76, 77], Long-User-Central-Window (LUCW) is considered, where the farther users are allocated to the central sub-bands of a THz transmission window, while the the users at shorter transmission ranges are provisioned with the edge sub-bands in a window. Such an interference management can also be employed within a drone network, where a drone (e.g., drone base station) serving different users can utilize dif- ferent sub-bands each several GHz wide [12], supporting capacity values in the order of several 100s of Gbps even under BM fading and MP fading as discussed earlier in Section 3. Nevertheless, interference in THz band com- munications usually occurs in dense scenarios [78, 79]. Thus, drone networks can be deployed strategically, e.g., via sharp pencil beam-forming, to mitigate the issue of the interference [6].

4.3 MAC and higher network layers

Similar to the physical layer-related issues, various challenges also emerge across the higher protocol stack layers. To start with the link layer, new medium access control (MAC) protocols are necessary to address the unique characteristics of the THz band as well as DSNs. Here, availability of the massive bandwidth annihilates the basic requirement for the communicating nodes to contest for the channel. Furthermore, THz signals with a miniature transmit duration also diminishes the chances of collision. Also, with the razor-sharp beams used in THz band transmission systems, MAC design for DSNs should facilitate receiver-initiated novel transmission schemes, so that the transmitter resources are not wasted when the intended receiver is unavailable. Novel MAC protocols managing the THz band communications for drone networks should also involve optimization of the packet size and error control techniques in an adaptive fashion. At the network layer, novel routing techniques should be devised to provision both the traditional active nodes (for relaying) together with new passive intelligent reflecting surfaces, which can relay the incoming THz signals towards the intended destination node [80]. Furthermore, novel metrics of routing are needed to be devised that captures the unique channel composition on the molecular level. These novel routing metrics can include the effect of the molecular composition on the distance and altitude-dependent THz bandwidth [12]. Across the transport layer, with THz band promising communication at 100s of Gbps or even up to Tbps, network congestion will drastically increase. This will give rise to issues across the transport layer pertaining to the flow/ congestion control, also ensuring end-to-end trans- port with reliability. For instance, it is expected that the conventional Transport Control Protocol (TCP) congestion control windowing will be revised to tackle the huge traffic demands of the THz band networks [81]. In addition to adapting MAC and higher layers to work efficiently with respect to the characteristics of the THz channel, the potential solutions should address the mobility of drones and DSN scenarios.

4.4 Security and privacy

Apart from the inherent advantages of the massive bandwidth and the huge rates THz band can offer drone networks with mobility [22]; security and privacy also comes up as a byproduct. Various unique characteristics of the THz waves highly influence the security and privacy [82]. With the unique THz absorption Spectra, secure wireless communications can coexist along with THz-based detection and imaging. Massive THz bandwidth favors anti-jamming approaches. However, the performance of such THz systems will be environment dependent, i.e., water vapor concentration levels. Huge attenuation at THz frequencies ensure link secrecy, while the razor-sharp THz beams favors covert communication, at the cost of rapid coordination of the communicating Tx and Rx beams. In regards to the privacy aspects, THz communications are hard to eavesdrop from a large distance due to the huge attenuation at THz frequencies near sea level. Meanwhile, a study shows that the THz signals can be intercepted by placing an object within the razor-sharp LOS THz beam for scattering the beam towards the eavesdropper [83]. For DSNs and drone networks at lower atmospheric altitudes (typically within a few 100s of meters above sea level), secure THz ultra-broadband communications can be established with the razor-sharp THz Tx-Rx beams pointing towards each other, leveraging the drone mobility, while the LOS beam scattering as shown in [83] will become difficult to realize for the eavesdroppers. This can be highly leveraged for sensitive applications, such as in military surveillance, border patrolling, etc. To summa- rize, we have studied several open issues and research directions towards realizing THz-enabled drone networks and DSNs from the physical layer to the higher network layers dealing with security and privacy issues. A similar study has recently been provided in [6], where the THz wireless systems have been defined based on seven features including: 1) Quasi-optical nature, 2) THz-based architectures, 3) Coexistence with the lower frequency bands, 4) Joint communication and sensing systems, 5) Physical layer strategies, 6) Spectrum access methods, and 7) Network optimization in real time. In the following, we overview the state-of-the-artartificial intelligence and machine leaning-based solutions to the problems related to the THz communications, particularly for the THz-enabled drone networks, in detail.

5. AI/ML BASED SOLUTIONS

The advent of Artificial Intelligence (AI) in the communications paradigm has been recently surged. Various tutorials and survey works have been published within the past few years on the AI/ML implementation for wireless communications [84, 85, 86, 87]. Among various recent studies, AI has been considered as the focus of 6G networks [88, 89, 90] for complementing the traditional methods. With the AI/ML technologies for Beyond 5G (B5G) systems, it will be possible to minimize/replace the existing manual network configuration management, as well as to ensure and deliver overall higher system performance with increased reliability. Moreover, communication networks will be able to adapt conveniently in real time based on the behavior of the users and the communication network. All in all, AI/ML will promise adaptive configuration and management of the communication networks by learning patterns and adapting to certain communication scenarios with flexibility, e.g., learning the communication traffic and planning in anticipa- tion [91]. A similar approach can be considered for the THz-enabled drone networks, where based on AI/ML techniques, a swarm of drones can adjust their respective positions optimally in order to route and maintain Tbps links among communicating drones. In what follows, we review some AI/ML-based approaches possibly implementable for THz-enabled drone networks for channel estimation, UM-MIMO, and Mobile edge computing.

5.1 Channel estimation

As discussed earlier in Section 1, first, the THz band is highly affected from absorption loss due to water vapor molecules in the atmosphere, contributing significantly to the total loss. Second, the spread loss is also massive at THz frequencies. Third, the THz band channels are non-stationary, particularly for mobile use cases i.e., hovering drones in our case, where both the Tx and Rx drones will be mobile. Hence, conventional assumptions of quasi-stationary or stationary channel models may not be applied to the THz band channels. More specifically, channel estimation over the THz band becomes more challenging in the drone scenarios under mobility, where precise Channel State Information (CSI) is needed, e.g., in beam-forming. Therefore, the traditional channel estimation methodologies are required to be revisited [52]. Overall, for reducing the complexity of the THz channel estimation, several techniques can be employed includ- ing: compressed sensing, fast channel tracking-based algorithms, etc. ML-based algorithms, in this context, can be employed for evaluating the THz band communication data by anticipating the THz signal loss in a certain unknown channel. Consequently, various AI or ML-based algorithms are applicable to the physical layer of the forth-coming 6G wireless networks for addressing the above-mentioned THz channel model and estimation [92, 93]. Supervised Learning (SL) [94] can aid in predicting THz shadowing and path loss. Moreover, SL can be employed for localization, channel estimation, interference management, etc. Employable SL models and algorithms are K-Nearest Neighbor (KNN), Support Vector Machine (SVM), feed-forward neural networks, and radial basis function neural networks, etc. Various challenges related to the THz channel modeling and estimation including multi-path tracking, interference mitigation, node clustering, optimized modulation, etc., can be tackled by using Unsupervised Learning (UL) techniques [95] such as: Fuzzy C-means, K-means, clustering algorithms, etc. Deep Learning (DL) (both SL and UL) can be employed in many aspects of channel modeling, such as for signal detection, and estimating Channel State Information (CSI). Techniques including Deep Neural Networks (DNNs), Recurrent Neural Networks (RNNs), Convolutional Neural Net- works (CNNs) can be anticipated as appropriate candidate DL algorithms [96]. Reinforcement Leaning (RL)[97] can be used for channel selection and tracking, iden- tification of radio bands, selection of modulation modes, etc. Appropriate RL models and techniques include Q-learning, fuzzy RL, etc. [98]. Finally, learning-based schemes for THz channel estimation is highly efficient particularly over higher dimensions [99]. As an example of deep kernel-based learning, a Gaussian process regression is studied in [100] for the channel estimation of a UM-MIMO multi-user system over the THz band (0.06-10 THz). It is to be noted here that the AI/ML techniques mentioned above are proposed for 6G wireless communications and networks at sea level in general. Therefore, there will be a need to tailor these AI/ML techniques specifically for the THz-enabled drone networks and DSNs keeping in view the intrinsic nature of the communicating drones i.e., mobility, energy-constrained resources etc.

5.2 UM-MIMO

In the realm of UM-MIMO, ML can be employed in various use cases. One instance is when an existing model is erroneous, and/or it is only a sparse approximate of the actual model. Such an instance can arise within the linear channel models, where the non-linearities induced by certain practical circumstances and hardware are neglected. Also, ML can be utilized for improving the solu- tions obtained using approximations of the linear mod- els. Another instance of ML-based solution in UM-MIMO is possible when the optimized solutions are computa- tionally expensive i.e., not feasible for the state-of-the-art hardware. Here, ML can be effectively utilized for finding suboptimal solutions having less complexity, with some obvious/acceptable lower performance. Examples in this context include channel estimation, maximum likelihood detection, etc. Moreover, optimum spectrum utilization in UM-MIMO can be made possible using machine learning techniques [22].

5.3 Mobile edge computing

Mobile Edge Computing (MEC) has recently emerged as a technique for 5G networks, in which cloud computing-like functionalities are processed at the edges of the cellular networks [101]. MEC can equip mobile devices, such as drones with constrained processing capabilities, to hand over their processing tasks to the nearest network's edge. Conversely, within a THz-enabled DSN, mobile user equipment can offload its computationally expensive jobs to the serving drones with MEC functionalities. Low latency systems for sporadic access such as cyber-physical communication systems (a.k.a., Tactile Internet) [102] require latencies within sub-ms for controlling hovering objects (drones in our case). Such alike systems will also be a requirement for evolved industry 4.0 applications [103]. It is predicted that the transport methods over the physical layer will be linked with edge computing such as MEC, or real-time cloud computing within the vicinity of the communication network. The main objective here is to deliver resources/solutions to the evolving IoT protocols comprising of a massive number of inter-connected devices with constrained energy and storage requirements such as in drone networks, as well with some latency requirements. For overcoming these constraints, global 6G research is moving towards distributed computation techniques (MEC here) [104, 105].

As an example, by em- ploying AI/ML using a DNN, initial extraction of the fea- tures can be obtained using drones, which are then re- layed at the network edges for the subsequent processing. Here, a THz band can be highly leveraged by relaying massive data to the network edges for MEC. In addition, novel energy efficient MAC protocols for THz-enabled drone networks and DSNs will be required for intelligently of- floading massive computations to the network edges, also considering drone mobility effects.

6. CONCLUSIONS

In this paper, we have summarized major characteristics of the various possible real-world applications of the THz-enabled drone networks and DSNs. After presenting the capacity potential of the THz-enabled drone networks via numerical results considering both ideal and realistic (fading) conditions of beam misalignment fading and multipath fading, we have discussed the major research challenges and directions for THz-enabled drone networks and DSNs, from physical layer channel estimation up to higher network layers, security/privacy issues. We outline and highlight AI/ML-based approaches as promising solutions. We proclaim that THz-enabled DSNs will be an integral part of the forthcoming 6G non-terrestrial networks. For development and validation of the THz band solutions across all layers, novel test beds for the real-world experiments are essential. Additionally, with the progression towards practical THz transceivers at sea level, novel THz transceivers for the drones will be required towards realizing 6G communications keeping in view the highly mobile nature of the drones, UM-MIMO given the battery limitations, and the sharp THz beams. Currently, such works are mostly limited to the near-THz transmission windows of 300-650 GHz, albeit, wireless systems at actual THz carrier frequencies (0.1 THz up to 10 THz) will be needed. Along with research directions mentioned in this paper, also further work is expected for standardizing and regulating the THz band.

ACKNOWLEDGEMENT

This work is supported in part by the Higher Education Commission (HEC), Government of Pakistan under the PhD scholarship program titled: HRDI Faculty Development of UESTPs Batch-V, SAP No: 50028241.

REFERENCES

[1] S. Hayat, E. Yanmaz, and R. Muzaffar, "Survey on Unmanned Aerial Vehicle Networks for Civil Applications: A Communications Viewpoint," *IEEE Communications Surveys Tutorials*, vol. 18, no. 4, pp. 2624–2661, 2016.

[2] E. Yanmaz, S. Yahyanejad, B. Rinner, H. Hellwagner, and C. Bettstetter, "Drone networks: Communications, coordination, and sensing," *Ad Hoc Networks*, vol. 68, pp. 1 – 15, 2018, advances in Wireless Communication and Networking for Cooperating Autonomous Systems. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1570870517301671

[3] G. Chmaj and H. Selvaraj, "Distributed Processing Applications for UAV/drones: A Survey," in *Progress in Systems Engineering*, H. Selvaraj, D. Zydek, and G. Chmaj, Eds. Cham: Springer International Publishing, 2015, pp. 449–454.

[4] M. Latva-aho, "Radio Access Networking Challenges Towards 2030," in *Proc. 1st International Telecommunication Union Workshop on Network 2030, New York*, Oct. 2018. [Online]. Available: https://www.itu.int/en/ITU-T/Workshops-and-Seminars/201810/Documents/Matt_Latvaaho_Presentation.pdf

[5] Z. Zhang, Y. Xiao, Z. Ma, M. Xiao, Z. Ding, X. Lei, G. K. Karagiannidis, and P. Fan, "6G Wireless Networks: Vision, Requirements, Architecture, and Key Technologies," *IEEE Vehicular Technology Magazine*, vol. 14, no. 3, pp. 28–41, 2019.

[6] C. Chaccour, M. N. Soorki, W. Saad, M. Bennis, P. Popovski, and M. Debbah, "Seven Defining Features of Terahertz (THz) Wireless systems: A Fellowship of Communication and Sensing," *arXiv*, 2021. [Online]. Available: https://arxiv.org/abs/2102.07668

[7] L. Bariah, L. Mohjazi, S. Muhaidat, P. C. Sofotasios, G. K. Kurt, H. Yanikomeroglu, and O. A. Dobre, "A Prospective Look: Key Enabling Technologies, Applications and Open Research Topics in 6G Networks," *IEEE Access*, vol. 8, pp. 174 792–174 820, 2020.

[8] A. A. Boulogeorgos, A. Alexiou, T. Merkle, C. Schubert, R. Elschner, A. Katsiotis, P. Stavrianos, D. Kritharidis, P. Chartsias, J. Kokkoniemi, M. Juntti, J. Lehtomaki, A. Teixeira, and F. Rodrigues, "Terahertz Technologies to Deliver Optical Network Quality of Experience in Wireless Systems Beyond 5G," *IEEE Communications Magazine*, vol. 56, no. 6, pp. 144–151, 2018.

[9] I. F. Akyildiz and J. M. Jornet, "Realizing Ultra-Massive MIMO (1024×1024) Communication in the (0.06-10) Terahertz band," *Nano Comm. Netw.*, vol. 8, pp. 46–54, 2016.

[10] I. F. Akyildiz, J. M. Jornet, and C. Han, "Terahertz Band: Next Frontier for Wireless Communications," *Physical Communication*, vol. 12, p. 16–32, Sep. 2014.

[11] A. Saeed, O. Gurbuz, and M. A. Akkas, "Terahertz Communications at Various Atmospheric

Altitudes," *Physical Communication*, vol. 41, p. 101113, 2020.

[12] A. Saeed, O. Gurbuz, A. O. Bicen, and M. A. Akkas, "Variable-Bandwidth Model and Capacity Analysis for Aerial Communications in the Terahertz Band," *IEEE Journal on Selected Areas in Communications*, pp. 1–1, 2021.

[13] V. Petrov, D. Moltchanov, M. Komar, A. Antonov, P. Kustarev, S. Rakheja, and Y. Koucheryavy, "Terahertz Band Intra-Chip Communications: Can Wireless Links Scale Modern x86 CPUs?" *IEEE Access*, vol. 5, pp. 6095–6109, 2017.

[14] J. M. Jornet and I. F. Akyildiz, "Channel Modeling and Capacity Analysis for Electromagnetic Wireless Nanonetworks in the Terahertz Band," *IEEE Trans. Wireless Communications*, vol. 10, no. 10, pp. 3211–3221, 2011. [Online]. Available: https://doi.org/10.1109/TWC.2011.081011.100545

[15] S. Mumtaz, J. Jornet, J. Aulin, W. Gerstacker, X. Dong, and b. ai, "Terahertz Communication for Vehicular Networks," *IEEE Transactions on Vehicular Technology*, vol. 66, pp. 5617–5625, 07 2017.

[16] S. Krishna Moorthy and Z. Guan, "Beam Learning in Mmwave/THz-band Drone Networks Under In-Flight Mobility Uncertainties," *IEEE Transactions on Mobile Computing*, pp. 1–1, 2020.

[17] Z. Guan and T. Kulkarni, "On the Effects of Mobility Uncertainties on Wireless Communications between Flying Drones in the mmWave/THz Bands," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WK- SHPS)*, 2019, pp. 768–773.

[18] X. Wang, P. Wang, M. Ding, Z. Lin, F. Lin, B. Vucetic, and L. Hanzo, "Performance Analysis of Terahertz Unmanned Aerial Vehicular Networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 16 330–16 335, 2020.

[19] R. Mendrzik, D. Cabric, and G. Bauch, "Error Bounds for Terahertz MIMO Positioning of Swarm UAVs for Distributed Sensing," in *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2018, pp. 1–6.

[20] G. Ding, Q. Wu, L. Zhang, Y. Lin, T. A. Tsiftsis, and Y. Yao, "An Amateur Drone Surveillance System Based on the Cognitive Internet of Things," *IEEE Communications Magazine*, vol. 56, no. 1, pp. 29–35, 2018.

[21] Z. Sun, P. Wang, M. C. Vuran, M. A. Al-Rodhaan, A. M. Al-Dhelaan, and I. F. Akyildiz, "BorderSense: Border Patrol through Advanced Wireless Sensor Networks," *Ad Hoc Networks*, vol. 9, no. 3, pp. 468–477, 2011.

[22] H. Sarieddeen, N. Saeed, T. Y. Al-Naffouri, and M. Alouini, "Next Generation Terahertz Communications: A Rendezvous of Sensing, Imaging, and Localization," *IEEE Communications Magazine*, vol. 58, no. 5, pp. 69–75, 2020 .

[23] S. Ergun and S. Sonmez, "Terahertz Technology For Military Applications," *Journal of Military and Information Science*, vol. 3, 01 2015.

[24] J. Federici, B. Schulkin, F. Huang, D. Gary, R. Barat, F. Oliveira, and D. Zimdars, "Thz Imaging and Sensing for Security Applications - Explosives, Weapons and Drugs," *Semiconductor Science and Technology*, vol. 20, p. S266, 06 2005.

[25] M. Erdelj, E. Natalizio, K. R. Chowdhury, and I. F. Akyildiz, "Help from the Sky: Leveraging UAVs for Disaster Management," *IEEE Pervasive Computing*, vol. 16, no. 1, pp. 24–32, 2017.

[26] T. Long, M. Ozger, O. Cetinkaya, and O. B. Akan, "Energy Neutral Internet of Drones," *IEEE Communications Magazine*, vol. 56, no. 1, pp. 22–28, 2018.

[27] N. Shimizu, H.-J. Song, Y. Kado, T. Furuta, A. Wakatsuki, and Y. Muramoto, "Gas Detection Using Terahertz Waves," vol. 7, 03 2009.

[28] T. S. Rappaport, Y. Xing, O. Kanhere, S. Ju, A. Madanayake, S. Mandal, A. Alkhateeb, and G. C. Trichopoulos, "Wireless Communications and Applications Above 100 GHz: Opportunities and Challenges for 6G and Beyond," *IEEE Access*, vol. 7, pp. 78 729–78 757, 2019.

[29] A. Fotouhi, M. Ding, and M. Hassan, "Flying Drone Base Stations for Macro Hotspots," *IEEE Access*, vol. 6, pp. 19 530–19 539, 2018.

[30] X. Wang, H. Zhang, Y. Tian, and V. C. M. Leung, "Modeling and Analysis of Aerial Base Station-Assisted Cellular Networks in Finite Areas Under LoS and NLoS Propagation," *IEEE Transactions on Wireless Communications*, vol. 17, no. 10, pp. 6985–7000, 2018.

[31] R. Singh and D. Sicker, "Reliable THz Communications for Outdoor based Applications- Use Cases and Methods," in *2020 IEEE 17th Annual Consumer Communications Networking Conference (CCNC)*, 2020, pp. 1–4.

[32] S. Garcia Sanchez, S. Mohanti, D. Jaisinghani, and K. R. Chowdhury, "Millimeter-wave Base Stations in the Sky: An Experimental Study of UAV-to-Ground Communications," *IEEE Transactions on Mobile Computing*, pp. 1–1, 2020.

[33] 3GPP, "Tsg.90-e plenaries," *Release 17*, December 12 2020.

[34] J. Angjo, I. Shayea, M. Ergen, H. Mohamad, A. Alhammadi, and Y. I. Daradkeh, "Handover Management of Drones in Future Mobile Networks: 6G Technologies," *IEEE Access*, vol. 9, pp. 12 803–12 823, 2021.

[35] W. Saad, M. Bennis, and M. Chen, "A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems," *IEEE Network*, vol. 34, no. 3, pp. 134–142, 2020.

[36] I. Bor-Yaliniz and H. Yanikomeroglu, "The New Frontier in RAN Heterogeneity: Multi-Tier Drone-Cells," *IEEE Communications Magazine*, vol. 54, no. 11, pp. 48–55, 2016.

[37] M. Mirahsan, R. Schoenen, and H. Yanikomeroglu, "HetHetNets: Heterogeneous Traffic Distribution in Heterogeneous Wireless Cellular Networks," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 10, pp. 2252–2265, 2015.

[38] A. A. Boulogeorgos, E. N. Papasotiriou, and A. Alexiou, "Analytical performance assessment of thz wireless systems," *IEEE Access*, vol. 7, pp. 11 436–11 453, 2019.

[39] S.A. Clough , M.W. Shephard , E.J. Mlawer , J.S. Delamere , M.J. Iacono , K. Cady-Pereira , S. Boukabara and P.D. Brown, "Atmospheric Radiative Transfer Modeling: a Summary of the AER Codes," *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 91, no. 2, pp. 233–244, 2005. [On-line]. Available: http://www.sciencedirect.com/science/article/pii/S0022407304002158

[40] S. A. Clough, M. J. Iacono, and J.-L. Moncet, "Line-by-Line Calculations of Atmospheric Fluxes and Cooling Rates: Application to Water Vapor," *Journal of Geophysical Research: Atmospheres*, vol. 97, no. D14, pp. 15 761–15 785, 1992. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/92JD01419

[41] P. Zhou, X. Fang, Y. Fang, R. He, Y. Long, and G. Huang, "Beam Management and Self-Healing for mmWave UAV Mesh Networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 1718–1732, 2019.

[42] M. T. Dabiri, H. Safi, S. Parsaeefard, and W. Saad, "Analytical Channel Models for Millimeter Wave UAV Networks Under Hovering Fluctuations," *IEEE Transactions on Wireless Communications*, vol. 19, no. 4, pp. 2868–2883, 2020.

[43] A. A. Farid and S. Hranilovic, "Outage Capacity Optimization for Free-Space Optical Links With Pointing Errors," *Journal of Lightwave Technology*, vol. 25, no. 7, pp. 1702–1710, 2007.

[44] M. D. Yacoub, "The α- μ distribution: A Physical Fading Model for the Stacy Distribution," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 1, pp. 27–34, 2007.

[45] M. Stojanovic and J. Preisig, "Underwater acoustic communication channels: Propagation models and statistical characterization," *IEEE Communications Magazine*, vol. 47, no. 1, pp. 84–89, January 2009.

[46] A. Fotouhi, M. Ding, and M. Hassan, "Flying Drone Base Stations for Macro Hotspots," *IEEE Access*, vol. 6, pp. 19 530–19 539, 2018.

[47] R. Singh and D. Sicker, "Beyond 5G: THz Spectrum Futures and Implications for Wireless Communication," International Telecommunications Society (ITS), 30th European Regional ITS Conference, Helsinki 2019 205213, June 2019. [Online]. Available: https://ideas.repec.org/p/zbw/itse19/ 205213.html

[48] G. Castellanos, M. Deruyck, L. Martens, and W. Joseph, "Performance Evaluation of Direct-Link Backhaul for UAV-Aided Emergency Networks," *Sensors*, vol. 19, no. 15, 2019.

[49] A. Goldsmith, *Wireless Communications*. New York, NY, USA: Cambridge University Press, 2005.

[50] Y. Xing and T. S. Rappaport, "Terahertz Wireless Communications: Co-Sharing for Terrestrial and Satellite Systems Above 100 GHz," *arXiv*, 2021. [Online]. Available: https://arxiv.org/abs/2103.00604

[51] C. Han and I. F. Akyildiz, "Distance-Aware Multi-Carrier (DAMC) Modulation in Terahertz Band Communication," in *2014 IEEE International Conference on Communications (ICC)*, 2014, pp. 5461–5467.

[52] H. Sarieddeen, M.-S. Alouini, and T. Y. Al-Naffouri, "An Overview of Signal Processing Techniques for Terahertz Communications," *arXiv*, 2020. [Online]. Available: https://arxiv.org/abs/2005.13176

[53] H. Elayan, O. Amin, B. Shihada, R. M. Shubair, and M. Alouini, "Terahertz Band: The Last Piece of RF Spectrum Puzzle for Communication Systems," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 1–32, 2020.

[54] R. C. Daniels and R. W. Heath, "60 GHz Wireless Communications: Emerging Requirements and Design Recommendations," *IEEE Vehicular Technology Magazine*, vol. 2, no. 3, pp. 41–50, 2007.

[55] 3GPP, "Study on New Radio Access Technology (Release 14)," *3GPP TR 38.801*, 2017.

[56] Y. Medjahdi, S. Traverso, R. Gerzaguet, H. Shaïek, R. Zayani, D. Demmer, R. Zakaria, J. Doré, M. Ben Mabrouk, D. Le Ruyet, Y. Louët, and D. Roviras, "On the Road to 5G: Comparative Study of Physi- cal Layer in MTC Context," *IEEE Access*, vol. 5, pp. 26 556–26 581, 2017.

[57] C. Han, A. O. Bicen, and I. F. Akyildiz, "Multi-Wideband Waveform Design for Distance-Adaptive Wireless Communications in the Terahertz Band," *IEEE Transactions on Signal Processing*, vol. 64, no. 4, pp. 910–922, Feb 2016.

[58] J. M. Jornet and I. F. Akyildiz, "Femtosecond-Long Pulse-Based Modulation for Terahertz Band Communication in Nanonetworks," *IEEE Transactions on Communications*, vol. 62, no. 5, pp. 1742–1754, 2014.

[59] Z. Hossain and J. M. Jornet, "Hierarchical Bandwidth Modulation for Ultra-Broadband Terahertz Communications," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, 2019, pp. 1–7.

[60] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A Survey of Non-Orthogonal Multiple Access for 5G," *IEEE Communications Surveys Tutorials*, vol. 20, no. 3, pp. 2294–2323, 2018.

[61] X. Zhang, C. Han, and X. Wang, "Joint Beamforming-Power-Bandwidth Allocation in Terahertz NOMA Networks," in *2019 16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, 2019, pp. 1–9.

[62] A. A. Boulogeorgos, E. N. Papasotiriou, and A. Alexiou, "A Distance and Bandwidth Dependent Adaptive Modulation Scheme for THz Communications," in *2018 IEEE 19th International Workshop on Sig- nal Processing Advances in Wireless Communica- tions (SPAWC)*, 2018, pp. 1–5.

[63] I. Akyildiz and J. Jornet, "Realizing Ultra-Massive MIMO (1024 x 1024) Communication in the (0.06-10) Terahertz Band," *Nano Communication Networks*, vol. 8, pp. 46–54, 6 2016.

[64] P. Chandhar and E. G. Larsson, "Massive MIMO for Connectivity With Drones: Case Studies and Future Directions," *IEEE Access*, vol. 7, pp. 94 676–94 691, 2019.

[65] A. Liao, Z. Gao, D. Wang, H. Wang, H. Yin, D. W. K. Ng, and M.-S. Alouini, "Terahertz Ultra-Massive MIMO-Based Aeronautical Communications in Space-Air-Ground Integrated networks," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 6, pp. 1741–1767, 2021.

[66] O. Tsilipakos, A. C. Tasolamprou, A. Pitilakis, F. Liu, X. Wang, M. S. Mirmoosa, D. C. Tzarouchis, S. Abadal, H. Taghvaee, C. Liaskos, A. Tsioliaridou, J. Georgiou, A. Cabellos-Aparicio, E. AlarcÃ³n, S. Ioannidis, A. Pitsillides, I. F. Akyildiz, N. V. Kantartzis, E. N. Economou, C. M. Soukoulis, M. Kafesaki, and S. Tretyakov, "Toward Intelligent Metasurfaces: The Progress from Globally Tunable Metasurfaces to Software-Defined Metasurfaces with an Embedded Network of Controllers," *Advanced Optical Materials*, vol. 8, no. 17, p. 2000783, 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/adom.202000783

[67] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M. S. Alouini, and R. Zhang, "Wireless Communications Through Reconfigurable Intelligent Surfaces," *IEEE Access*, vol. 7, pp. 116 753–116 773, 2019.

[68] M. Di Renzo, M. Debbah, D.-T. Phan-Huy, A. Zappone, M. S. Alouini, C. Yuen, V. Sciancalepore, G. Alexandropoulos, J. Hoydis, H. Gacanin, J. D. Rosny, A. Bounceur, G. Lerosey, and M. Fink, "Smart Radio Environments Empowered by Reconfigurable AI Meta-Surfaces: An Idea Whose Time Has Come," *EURASIP Journal on Wireless Communications and Networking*, no. 129, 2019.

[69] R. Alghamdi, R. Alhadrami, D. Alhothali, H. Almorad, A. Faisal, S. Helal, R. Shalabi, R. Asfour, N. Hammad, A. Shams, N. Saeed, H. Dahrouj, T. Y. Al-Naffouri, and M. S. Alouini, "Intelligent Surfaces for 6G Wireless Networks: A Survey of Optimization and Performance Analysis Techniques," *IEEE Access*, vol. 8, pp. 202 795–202 818, 2020.

[70] C. Chaccour, M. N. Soorki, W. Saad, M. Bennis, and P. Popovski, "Risk-Based Optimization of Virtual Reality over Terahertz Reconfigurable Intelligent Surfaces," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.

[71] A.-A. A. Boulogeorgos and A. Alexiou, "Coverage Analysis of Reconfigurable Intelligent Surface Assisted THz Wireless Systems," *IEEE Open Journal of Vehicular Technology*, vol. 2, pp. 94–110, 2021.

[72] X. Ma, Z. Chen, W. Chen, Y. Chi, Z. Li, C. Han, and Q. Wen, "Intelligent Reflecting Surface Enhanced Indoor Terahertz Communication Systems," *Nano Communication Networks*, vol. 24, p. 100284, 2020.

[73] N. Abuzainab, M. Alrabeiah, A. Alkhateeb, and Y. E. Sagduyu, "Deep Learning for THz Drones with Flying Intelligent Surfaces: Beam and Handoff Prediction," *arXiv*, 2021. [Online]. Available: https: //arxiv.org/abs/2102.11222

[74] M. Matinmikko-Blue, S. Yrj lä, and P. Ahokangas, "Spectrum Management in the 6G Era: The Role of Regulation and Spectrum Sharing," in 2020 2nd 6G Wireless Summit (6G SUMMIT), 2020, pp. 1–5.

[75] B. Aazhang, P. Ahokangas, H. Alves, M.-S. Alouini, J. Beek, H. Benn, M. Bennis, J. Belfiore, E. Strinati, F. Chen, K. Chang, F. Clazzer, S. Dizit, D. Kwon, M. Giordiani, W. Haselmayr, J. Haapola, E. Hardouin, E. Harjula, and P. Zhu, *Key Drivers and Research Challenges for 6G Ubiquitous Wireless Intelligence (White Paper)*, 09 2019.

[76] C. Han and I. F. Akyildiz, "Distance-Aware Bandwidth-Adaptive Resource Allocation for Wireless Systems in the Terahertz Band," *IEEE Transactions on Terahertz Science and Technology*, vol. 6, no. 4, pp. 541–553, 2016.

[77] M. Yu, A. Tang, X. Wang, and C. Han, "Joint Scheduling and Power Allocation for 6G Terahertz Mesh Networks," in *2020 International Conference on Computing, Networking and Communications (ICNC)*, 2020, pp. 631–635.

[78] C. Chaccour, M. N. Soorki, W. Saad, M. Bennis, and P. Popovski, "Can Terahertz Provide High-Rate Reliable Low Latency Communications for Wireless VR?" *arXiv*, 2021. [Online]. Available: https://arxiv.org/abs/2005.00536

[79] V. Petrov, D. Moltchanov, and Y. Koucheryavy, "Interference and SINR in Dense Terahertz Networks," in *2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall)*, 2015, pp. 1–5.

[80] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent Reflecting Surface Aided Wireless Communications: A Tutorial," *IEEE Transactions on Communications*, pp. 1–1, 2021.

[81] NetWorld2020, "Strategic Research and Innovation Agenda 2021-27: Smart Networks in the Context of NGI," *European Technology Platform NetWorld2020*, 2018.

[82] R. Singh and D. Sicker, "Thz Communications - A Boon and/or Bane for Security, Privacy, and National Security," *SSRN Electronic Journal*, 01 2020.

[83] J. Ma, R. Shrestha, J. Adelberg, C.-Y. Yeh, Z. Hossain, E. Knightly, J. Jornet, and D. Mittleman, "Security and Eavesdropping in Terahertz Wireless Links," *Nature*, vol. 563, 11 2018.

[84] X. Wang, Y. Han, V. C. M. Leung, D. Niyato, X. Yan, and X. Chen, "Convergence of Edge Computing and Deep Learning: A Comprehensive Survey," *IEEE Communications Surveys Tutorials*, vol. 22, no. 2, pp. 869–904, 2020.

[85] N. C. Luong, D. T. Hoang, S. Gong, D. Niyato, P. Wang, E. Liang, and D. I. Kim, "Applications of Deep Reinforcement Learning in Communications and Networking: A Survey," *IEEE Communications Surveys Tutorials,* vol. 21, no. 4, pp. 3133–3174, 2019.

[86] Y. Sun, M. Peng, Y. Zhou, Y. Huang, and S. Mao, "Application of Machine Learning in Wireless Networks: Key Techniques and Open Issues," *IEEE Communications Surveys Tutorials*, vol. 21, no. 4, pp. 3072–3108, 2019.

[87] J. Jagannath, N. Polosky, A. Jagannath, F. Restuccia, and T. Melodia, "Machine Learning for Wireless Communications in the Internet of Things: A Comprehensive Survey," *Ad Hoc Networks*, vol. 93, p. 101913, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1570870519300812

[88] L. Zhang, Y. Liang, and D. Niyato, "6G Visions: Mobile Ultra-Broadband, Super Internet-of-Things, and Artificial Intelligence," *China Communications*, vol. 16, no. 8, pp. 1–14, 2019.

[89] G. Gui, M. Liu, F. Tang, N. Kato, and F. Adachi, "6G: Opening new horizons for integration of comfort, security, and intelligence," *IEEE Wireless Communications*, vol. 27, no. 5, pp. 126–132, 2020.

[90] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y. A. Zhang, "The Roadmap to 6G: AI Empowered Wireless Networks," *IEEE Communications Magazine*, vol. 57, no. 8, pp. 84–90, 2019.

[91] C.-X. Wang, M. D. Renzo, S. Stanczak, S. Wang, and E. G. Larsson, "Artificial intelligence enabled wireless networking for 5G and beyond: Recent advances and future challenges," *IEEE Wireless Communications*, vol. 27, no. 1, pp. 16–23, 2020.

[92] J. Du, C. Jiang, J. Wang, Y. Ren, and M. Debbah, "Machine Learning for 6G Wireless Networks: Carrying Forward Enhanced Bandwidth, Massive Access, and Ultrareliable/Low-Latency Service," *IEEE Vehicular Technology Magazine*, vol. 15, no. 4, pp. 122–134, 2020.

[93] A. Zappone, M. Di Renzo, and M. Debbah, "Wireless Networks design in the Era of Deep Learning: Model-based, AI-Based, or Both?" *IEEE Transactions on Communications*, vol. 67, no. 10, pp. 7331–7376, 2019.

[94] M. E. Morocho-Cayamcela, H. Lee, and W. Lim, "Machine Learning for 5G/B5G Mobile and Wireless Communications: Potential, Limitations, and Future Directions," *IEEE Access*, vol. 7, pp. 137 184–137 206, 2019.

[95] R. Nikbakht, A. Jonsson, and A. Lozano, "Unsupervised Learning for Parametric Optimization," *IEEE Communications Letters,* vol. 25, no. 3, pp. 678–681, 2021.

[96] L. Dai, R. Jiao, F. Adachi, H. V. Poor, and L. Hanzo, "Deep Learning for Wireless Communications: An Emerging Interdisciplinary Paradigm," *IEEE Wireless Communications,* vol. 27, no. 4, pp. 133–139, 2020.

[97] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. MIT Press, 2017.

[98] Y. S. Nasir and D. Guo, "Multi-Agent Deep Reinforcement Learning for Dynamic Power Allocation in Wireless Networks," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 10, pp. 2239–2250, 2019.

[99] H. He, R. Wang, W. Jin, S. Jin, C.-K. Wen, and G. Y. Li, "Beamspace Channel Estimation for Wideband Millimeter-Wave MIMO: A Model-Driven Unsupervised Learning Approach," *arXiv*, 2021. [Online]. Available: https://arxiv.org/abs/2006.16628

[100] S. Nie and I. F. Akyildiz, "Deep Kernel Learning-Based Channel Estimation in Ultra-Massive MIMO Communications at 0.06-10 thz," in *2019 IEEE Globecom Workshops (GC Wkshps)*, 2019, pp. 1–6.

[101] A. A. Al-Habob, A. Ibrahim, O. A. Dobre, and A. G. Armada, "Collision-Free Sequential Task Offloading for Mobile Edge Computing," *IEEE Communications Letters*, vol. 24, no. 1, pp. 71–75, 2020.

[102] G. Fettweis, "The Tactile Internet: Applications and Challenges," *Vehicular Technology Magazine, IEEE*, vol. 9, pp. 64–70, 03 2014.

[103] H. Lasi, P. Fettke, H.-G. Kemper, T. Feld, and M. Hoffmann, "Industry 4.0," *Business and Information Systems Engineering*, vol. 6, pp. 239–242, 08 2014.

[104] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. A. Zhang, "The Roadmap to 6G: AI Empowered Wireless Networks," *IEEE Communications Magazine*, vol. 57, no. 8, pp. 84–90, 2019.

[105] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A Survey on Mobile Edge Computing: The Communication Perspective," *IEEE Communications Surveys Tutorials*, vol. 19, no. 4, pp. 2322–2358, 2017.

AUTHORS

Dr. Akhtar Saeed received his B.E. degree in telecommunications engineering at U.I.T Hamdard University in 2010 and M.E. degree in telecommunications engineering at NED University of Engineering and Technology in 2014. He received his Ph.D. degree in electronics engineering from Sabanci University in 2021 under the supervision of Dr. Ozgur Gurbuz and co-supervision of Dr. Mustafa Alper Akkaş. His research interests include terahertz communications at various atmospheric altitudes, millimeter wave communications, wireless sensor networks and full duplex communications.

Dr. Ozgur Gurbuz received her B.S. and M.S. degrees in electrical and electronics engineering at Middle East Technical University, in 1992 and 1995, respectively. She received her Ph.D. degree in electrical and computer engineering from Georgia Institute of Technology in 2000. From 2000 until 2002 she worked as a researcher and systems/algorithms engineer for Cisco Systems, in Wireless Access and Wireless Networking Business Units. As of September 2002, Dr. Gurbuz joined the Faculty of Engineering and Natural Sciences at Sabanci University, where she is serving as a professor. Her research interests remain in the field of wireless communications and networks, specifically design of link and higher layer network algorithms/protocols for emerging physical layer techniques, including full-duplex communication, cooperative communication, MIMO, smart antennas. Recently, she has been working on full-duplex communications, terahertz communications and applications of machine learning in wireless communications/networks.

Dr. Mustafa Alper Akkaş received his B.S. degree in electrical and electronics engineering from Erciyes University in 2006. He received his Ph.D. degree in the Department of Electrical and Electronics Engineering at Ege University in 2014. Currently, he is an associate professor at the Department of Computer Engineering at Bolu Abant Izzet Baysal University. His research interests include terahertz-band communication networks, channel modeling, wireless underground communication networks, internet of things and intra-body wireless nanosensor networks. He was a visiting student at Georgia Institute of Technology, in the Broadband Wireless Networking Lab under the supervision of Prof. Ian F. Akyildiz from September 2011 to May 2012.

Dr. Ahmet Ozan Bicen received B.Sc. degree in electrical and electronics engineering from Middle East Technical University, Ankara, Turkey, in 2010, M.Sc. degree in electrical and electronics engineering from Koc University, Istanbul, Turkey, in 2012, and Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2016. He was a postdoctoral researcher in the same department in between 2016-2018. He was an Assistant Professor in the Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul, Turkey, in between 2019-2021. His current research interests include statistical signal processing, molecular communication, and computational physiology.

M-ARY QUADRATURE AMPLITUDE MODULATION ORDER OPTIMIZATION FOR TERAHERTZ WIRELESS COMMUNICATIONS OVER DISPERSIVE CHANNELS

Karl Strecker[1], Sabit Ekin[1], and John O'Hara[1]
[1]Department of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK, 74078 USA

NOTE: Corresponding author: John O'Hara, oharaj@okstate.edu

Abstract – *Highly accurate atmospheric models, based on molecular resonance information contained within the HITRAN database, were used to simulate the propagation of high capacity single-carrier quadrature amplitude modulated signals through the atmosphere for various modulation orders. For high-bandwidth signals such as those considered in this work, group velocity dispersion caused by atmospheric gases distorts the modulated waveform, which may produce bit errors. This leads to stricter Signal-To-Noise Ratio requirements for error-free operation, and this effect is more pronounced in high-order modulation schemes. At the same time, high-order modulation schemes are more spectrally efficient, which reduces the bandwidth required to maintain a given data rate, and thus reduces the total group velocity dispersion in the link, resulting in less distortion and better performance. Our work with M-ary quadrature amplitude modulated signals shows that optimal selection of modulation order can minimize these conflicting effects, resulting in decreased error rate, and reducing the performance requirements placed on any equalizers, other dispersion-compensating technologies, or signal processing hardware.*

Keywords – Atmospheric modeling, bit error rate, chromatic dispersion, millimeter wave communications, quadrature amplitude modulation, terahertz communications

1. INTRODUCTION

Wireless data rates have risen dramatically over the last decade, and are projected to continue to do so over the decade to come [1, 2]. This growth has been fueled by demand, created by consumer expectations as well as new technologies such as virtual reality, high-definition video streaming, and (most significantly) the Internet of Things (IoT) [3, 4]. This growth has been enabled by the development of devices capable of operating at progressively higher frequencies and bandwidths. Wireless systems operating at several gigahertz are commercially available off-the-shelf, and networks operating at several tens of gigahertz (millimeter wave) are just on the verge of becoming so. The inevitable next step is systems operating at sub-millimeter wavelengths, that is, hundreds of gigahertz [5]. This is frequently recognized as the beginning of the terahertz communication bands. These bands have been slow in development for many years, in part due to the challenge of atmospheric absorption and in part due to the technological difficulties arising from the fact that few devices are naturally active in these frequencies.

However, the so-called "terahertz gap" is beginning to close [6]. Recent progress in terahertz devices has resulted in hardware not only capable of producing and processing these high-speed signals, but also powerful enough to overcome the atmospheric attenuation, which is much more severe than at microwave frequencies. Over the last decade, several prototype terahertz communication systems have been demonstrated, operating in the hundreds of gigahertz, achieving communications over multi-kilometer distances.

For example, in 2010, Hirata *et al.* demonstrated a wireless link operating at 120 GHz, using Binary Phase Shift Keying (BPSK) that achieved an error-free data rate of 10 Gb/s over 5 km [7]. In 2013, Takahashi *et al.* also demonstrated a 10-Gb/s, error free link at 120 GHz, using Quadrature Phase Shift Keying (QPSK, or 4-QAM), over a distance of 170 m [8]. However, their calculations indicated the link could conceivably span up to 2 km. The same year, another wireless link was demonstrated, this time at 140 GHz, using 16-QAM to achieve 10 Gb/s over 1.5 km, with an error rate of 10^{-6} [9].

In 2017, another communication link centered at 94 GHz, using 8-QAM, achieved a data rate of 54 Gb/s, with an error rate of 3.8×10^{-3}, over 2.5 km [10]. Also in 2017, Kallfass *et al.* presented a review of their experimental work with point-to-point millimeter wave links which included an E-band link (between 60 and 90 GHz, carrier frequency not specified) and a 240 GHz link [11]. The E-band link used QPSK, 8-QAM, and 16-QAMs, and achieved data rates in the range of 4 Gb/s up to 21 Gb/s, over ranges between 4.1 km and 36.7 km, under various weather con- ditions with error rates below 4.8×10^{-3}. The 240 GHz link used QPSK modulation, and achieved 64 Gb/s over 0.85 km, with an error rate of 7.9×10^{-5}. Many different link configurations were investigated in the review, and the reader is referred to the work of Kalfass *et al.* for more detailed information [11].

Finally, Wu *et al.* also demonstrated a long-distance wireless communication system at 140 GHz in 2017, which spanned 21 km and used 16-QAM to achieve 5 Gb/s with effectively error-free operation (a bit-error rate below

10^{-12}) [12]. The authors also estimated that their system could extend to span even farther ranges with the use of more advanced error-correction codes.

Collectively, these demonstrations indicate that long-range terahertz communication links are not only possible, but will likely be implemented commercially in the foreseeable future as the technology progresses. This technology would provide many benefits, since there are many situations in which the ability to rapidly establish a directive, wireless point-to-point link with a capacity of tens of Gb/s would be highly attractive, including temporary installments during disaster recovery or wartime environments, at locations where trenching fiber is prohibitively expensive or time-consuming, or as a re- placement to upgrade microwave point-to-point backhaul links.

There are many design choices that must be considered when planning the construction of such a link [13, 14, 15, 16]. One notable known design choice is selecting the modulation scheme, since modulation type determines the shape of the temporal waveform, the hardware requirements (for example, the dynamic range of the front-end receiver), the resilience of the channel to interference, and the achievable throughput of the channel. By judicious selection of the modulation type, channel throughput can be maximized, and many research teams have investigated various algorithms and strategies for determining the optimal modulations for both microwave and terahertz wireless links [17, 18, 19].

Many modulation schemes are possible, and all carry their own benefits and drawbacks. However, the prototype terahertz links in the demonstrations listed earlier em- ploy various orders of quadrature amplitude modulation, collectively known as M-QAM schemes, including Binary Phase Shift Keying (BPSK, or 2-QAM), Quadrature Phase Shift Keying (QPSK, or 4-QAM), 8-QAM and 16-QAM. In an M-QAM scheme, binary data is encoded as communi- cation symbols, distinct combinations of amplitude and phase of the carrier wave, each of which represent one or more bits of data. The modulation order M specifies how many such combinations of amplitude and phase are recognized by the receiver, and $\log_2(M)$ bits of data are carried by each symbol.

As the modulation order of the communication system is increased, each symbol transition carries more information, which consequently increases the spectral efficiency of the link. Spectral efficiency is a measure of how many bits of data are transferred per unit of bandwidth utilized by the communication system, typically given in units of $(\frac{\text{bits}}{s})/\text{Hz}$. While the spectral efficiency realized in a physical communication system depends on many factors (such as the coding scheme, Signal-To-Noise Ratio (SNR), and fading characteristics of the channel), the theoretical maximum spectral efficiency of an M-QAM scheme is ultimately given by, and scales with modulation order according to, log2(M) [20].

In general, this increase in spectral efficiency makes higher order modulations the most attractive, due to the fact that more information can be sent within in a given bandwidth or, conversely, that the system requires less bandwidth to maintain a given data rate.

However, higher-order modulations are not always viable to use. When a system is constrained to operate below some fixed maximum power, the phase and amplitude of all communication symbols must fall within a finite region of the phase/amplitude plane that satisfies that power constraint. Increasing the number of communication symbols necessarily means that symbols must take on increasingly similar values of amplitude and/or phase, as more symbols have to be placed within the finite region satisfying the power constraint. When the receiver must differentiate between a large number of similar symbols with high resolution, injected noise can easily shift the amplitude and/or phase of the received waveform so that symbols are received in error, much more so than for a lower-order modulation scheme where symbol regions are larger and more widely spaced. As a result, higher-order M-QAM schemes have more stringent requirements on the minimum SNR allowed at the receiver for effective operation. Fig. 1 illustrates the increase in SNR required by higher-order M-QAM schemes in order to maintain a given bit error rate.

When designing a communication link, the modulation type is chosen so that an acceptable error rate is maintained under the worst-case SNR the link is designed to handle. In order to decrease the outage probability during times when the signal is strongly attenuated, and to increase the capacity when channel conditions are favorable, many communication systems employ optimization algorithms that actively select the order of the modulation scheme used [17]. These optimization routines switch between modulation orders as channel conditions vary, such that the resulting link is both more reliable (in terms of outage probability) and operates with a higher average capacity.

Terahertz links will, of course, likewise benefit from these type of optimization routines [18, 17], whether the bandwidth is occupied by a single link, or filled with a large number of subcarriers [21]. However, due to the huge bandwidths available for terahertz communication links, and the high frequencies at which they operate, the optimal modulation type will not be determined by SNR (that is, fading) alone. Our work indicates that the Group Velocity Dispersion (GVD) caused by molecular resonances in the atmosphere can result in counter-intuitive behavior over the lower terahertz bands, in which the severity of Inter-Symbol Interference (ISI) depends not only on band- width (as expected), but also on the modulation type used, even in the absence of noise.

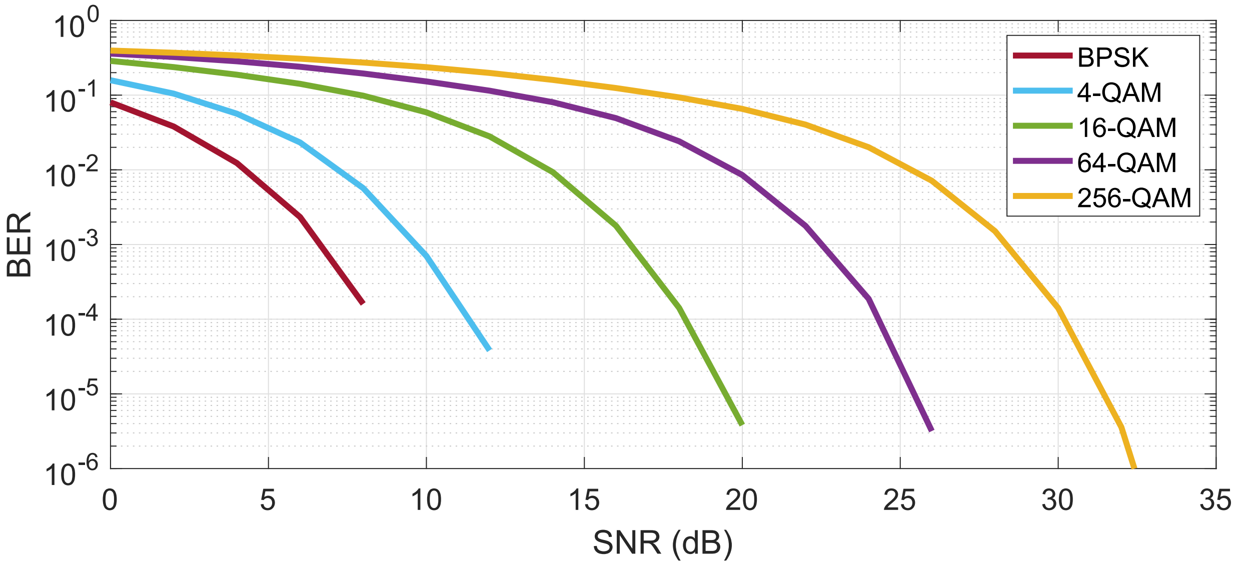


Fig. 1 – Bit error rate versus SNR "waterfall" plots for an M-QAM communication system, with M = 2, 4, 16, 64, and 256. Higher-order modulations have closer symbol spacing under equivalent power requirements, resulting in a higher SNR required for equivalent error performance to a lower order modulation, assuming the absence of group velocity dispersion.

2. METHODOLOGY

In order to quantitatively measure the impact of ISI caused by atmospheric GVD, bit error rate simulations were performed using a channel model founded upon an accurate understanding of atmospheric molecular resonances. It is from this atmospheric model that all the effects accounted for in this work were derived. Specifically, the channel considered in this study was a Linear Time-Invariant (LTI) channel with Additive White Gaussian Noise (AWGN) and no obstruction, multipath propagation, or Doppler effects. However, the transfer function of the atmosphere itself was modeled as variable over frequency in both absorption and refractive index, which gives rise to the behavior observed in our results. Even though our assumption of an LTI AWGN channel is much simpler than the environments usually encountered by wireless link designers at terahertz frequencies, the fact that our results arise from the properties of the atmosphere rather than complex and situation-specific channel effects make them applicable to a wide range of channels, including those significantly more complex that that presented here [22].

The atmospheric transfer function is described most generally as $H_a(\omega) = \alpha(\omega)\exp[-j\phi(\omega)]$, where $\alpha(\omega)$ and $\phi(\omega)$ are the frequency-dependent attenuation and phase shift imparted by the atmosphere, respectively, and $j = \sqrt{-1}$. This non-unity transfer function arises from the interaction of various atmospheric gas species with terahertz-frequency radiation. Most notable among these are water vapor and diatomic oxygen, which exhibit strong rotational and vibrational resonances within and above the terahertz bands. While the amplitude (absorption) term of $H_a(\omega)$ is most often discussed, the phase term $\phi(\omega)$ is equally important to propagation, and together these terms determine the complex index of refraction of the atmosphere. This complex index is obtained by a combination of Molecular Response Theory (MRT) [23] and continuum effects [24, 25], in which the broadened absorption lines of all the H_2O and O_2 molecular resonances from 0 to 5 THz are found by MRT, summed, then added to the continuum absorption. This has been shown to accurately model atmospheric behavior over the subterahertz bands, and accounts for the contribution of all relevant molecular resonances up to 5 THz.

In addition to the atmospheric effects, pulse shaping filters also shape the transmitted waveform, limit the bandwidth of the signal, and reduce ISI. In our simulations, a raised cosine filter $H_{rc}(\omega)$ with a roll-off factor of 1 was used, and incorporated into the channel model by applying it directly to the atmospheric transfer function in frequency domain, yielding a channel transfer function $H_c(\omega) = H_{rc}(\omega) \times H_a(\omega)$. The impulse response of the complex channel transfer function can then be derived as $h_c(t) = \mathcal{F}^{-1}[H_c(\omega)]$, where $\mathcal{F}^{-1}$ indicates the inverse Fourier transform.

Once the impulse response of the channel is known, a data vector containing complex valued communication symbols is generated. The symbols in the data stream occur with equal distribution, but the data stream is not completely random. Rather, it is generated such that *combinations* of symbols are also equally distributed, so every possible permutation of k symbols occurs an equal number of times for a specified k. This is necessary because the severity of ISI experienced by a communication symbol depends on the value and order of the neighboring symbols, not on the value of the symbol itself. This data stream is convolved with the channel impulse response, resulting in a sequence of non-ideal communication symbols in the time domain, which have experienced a nonlinear phase shift due to GVD in the atmosphere, resulting in ISI in the time domain. At this point, AWGN is added to the distorted symbol sequence, which is then demodulated and compared to the original symbol vector to determine the number of errors. This procedure is repeated until enough iterations have run to ensure the errors generated are true to the stochastic distribution of the noise. Averaging the error rates observed on each iteration gives the error rate of the link for that particular combination of distance, bandwidth, modulation type, and atmospheric properties.

This simulation process has the advantage of abstracting away much of the hardware, focusing on the mathematical, fundamental interactions between the data-carrying symbols and the atmospheric channel. Notably, it also defines parameters such as received SNR directly prior to demodulation, so that the results are applicable to a broad range of physical systems. For a much more in-depth description of the simulation process using a slightly different mathematical convention, readers are referred to our previous work [26], which is currently under consideration for publication at the time of this writing.

3. RESULTS AND DISCUSSION

A principal result taken from our simulations is that GVD can produce increasing SER in two opposing cases: when low bandwidth and high order modulation is employed and when high bandwidth and low order modulation is employed. Moreover, there exists an optimum tradeoff between modulation order and bandwidth that minimizes SER due to GVD for a particular desired data rate. In order to clarify this point, we have chosen to conduct our simulations such that data rate is held constant, whereas bandwidth and modulation order are variable. Further justification for this approach is offered later in the discussion. To elaborate on our results, the following relationships between modulation order and GVD were found:

For a high bandwidth link with low modulation order, achievable data rate is high, but the link suffers a high number of symbol errors because of the large frequency-dependent change in refractive index across the bandwidth (high GVD) causes severe ISI, large enough to push received symbols across the broadly-spaced decision boundaries used in low-order modulation types.

For a low bandwidth link with high order modulation, the achievable data rate is equally high, but the link still suffers a high number of symbol errors due to dispersion. This time, the errors are not because of a large frequency dependency in the narrow channel, but because decision boundaries are so tightly spaced on the constellation diagram that even the small amount of GVD exhibited by the channel is enough to push received symbol values across them, again causing ISI.

A compromise between these two extremes allows for high bit rates with higher dispersion tolerance when bandwidth and modulation order are properly balanced. Note that the remaining combinations of bandwidth and modulation order fare quite poorly: a high bandwidth link with high order modulation can have a very high capacity but suffers severe ISI due to simultaneously high dispersion and tight decision boundaries, while a low band-width link with low modulation order has greatly reduced data capacity, defeating the purpose of terahertz commu- nications. An example of the simulation results that led us to these conclusions is illustrated in Fig. 2, which shows the bit-error rate of a 60 Gb/s M-QAM communication link for five different modulation orders over 0 to 20 km.

It is important to note that atmospheric dispersion is a cumulative phenomenon, meaning the greater distance a signal propagates, the more dispersion accumulates and affects that signal. It is also important to note that no noise is added to the signal in Fig. 2, which allows us to confidently state that any change in error rates observed are due to GVD increasing with distance, and the that differences between the various curves are due to changes in modulation order (which affects both bandwidth and symbol spacing on the constellation diagram). Even in the absence of noise, increasing dispersion over distance will eventually cause some communication symbols to be misinterpreted by the receiver for all modulation types, resulting in a bit error rate that rises rapidly from insignificance to some finite value, often by several orders of magnitude in only a kilometer or two.

The point at which the error rate jumps from insignificance to a finite value is the uncompensated “dispersion limit” of that link, marked by vertical dashed lines in Fig. 2. Beyond this limit, the error rate of the link cannot be continuously improved by increasing the SNR, because dispersion is the dominant source of errors [26]. For the 60 Gb/s link shown in Fig. 2, BPSK has the lowest dispersion limit, meaning it is most severely affected by atmospheric temporal dispersion. 4-QAM, or QPSK, is next, followed by 256-QAM, 16-QAM, and finally 64-QAM. In other words, the dispersion limit increases with modulation order for most of the modulation orders simulated, meaning there is more robust operation as the bandwidth decreases.

The results and discussion presented so far may seem obvious and well-established. It is well-known that decreasing the bandwidth of a wireless link operating in a frequency-selective environment will increase the performance of the link by ”flattening” the fading profile of the channel, thereby reducing errors, ISI, and the complexity of signal processing. The atmosphere is a frequency-selective channel over the huge bandwidths available to terahertz communication links, so it may not initially seem surprising that as we decrease the bandwidth we also observe an improvement in error rate. However, closer inspection of the data reveal additional and unexpected behaviors that are not readily explained

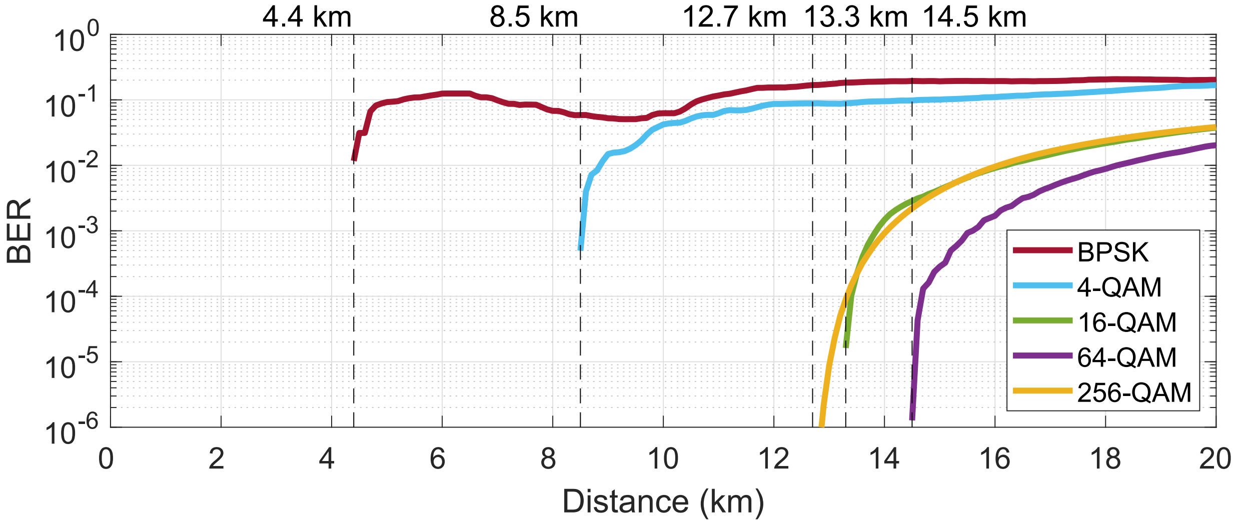


Fig. 2 – Bit error rate versus distance "reverse waterfall" plot for a noiseless 60 Gb/s link, centered at 250 GHz. Dashed vertical lines mark "dispersion limits," the distance at which uncompensated dispersion begins to deterministically cause bit errors, which cannot be overcome by increasing the SNR. Atmospheric conditions are water vapor density $\rho_{wv} = 10.37\ \mathrm{g/m^3}$ (60% relative humidity at 20 °C). The decrease in the error rate of the BPSK curve over 7 km to 10 km is a consequence of how atmospheric GVD shifts the received value of communication symbols. Because BPSK modulates only a single dimension in the complex symbol space, it experiences fewer errors over the 7 km to 10 km region, where dispersion tends to shift a majority of symbols orthagonally to the dimension of modulation. This effect is not strongly observed in higher-order modulations due to their use of the orthogonal (quadrature) dimension, though this effect also produces a slight dip at about 9 km for the 4-QAM link.

by the usual intuition about wireless systems. If frequency-selective fading was the driver of the increase in error rate, then we would expect to see the error rate improve with every decrease in bandwidth, but the exact opposite is observed for the transition from 64-QAM to 256-QAM. In fact, 256-QAM has a similar dispersion limit to 16-QAM, despite having *twice* the spectral efficiency, that is, half the bandwidth. The cause for this reversal is that the symbols in 256-QAM are so closely spaced that only a small amount of dispersion is enough to shift them across the decision boundaries and produce errors, even though the frequency-dependent fading due to the atmosphere is essentially flat across the bandwidth. In other words, in the presence of GVD, the decrease in symbol spacing outweighs the decrease in bandwidth due to spectral efficiency gains, resulting in more errors.

This demonstrates that GVD, not frequency-selective fading, is responsible for these errors. It is worth remembering that the results shown in Fig. 2 are for a single link, with no multipath interference, over an LTI channel with no noise added. The errors observed are solely due to the frequency dependent refractive index of the atmosphere. Consequently, these results show that in the terahertz and sub-terahertz bands, reducing bandwidth *does not necessarily* improve error rate performance because the shape of the waveform (that is, modulation type) also matters, due to the atmospheric interaction. This is a counter-intuitive result that is uniquely different from free-space microwave links.

While the noiseless case is instructive, it is not always rep- resentative of the real world. Fig. 3 shows the error per- formance of the 4-QAM and 16-QAM links of Fig. 2 in the presence of varying amounts of noise, as described in the figure caption. As expected, poor SNR impacts the error rate of the high-order 16-QAM link more severely than the 4-QAM link. It is significant to note that at the lower order 4-QAM link, the error rate can be improved in some cir- cumstances, even with poor SNR, by shifting to a higher or- der modulation, whenever the cost of degraded SNR performance is offset by reduced GVD in the more spectrally efficient modulation.

For example, examine the 4-QAM link operating at 8 kilometers with an SNR of 20 dB, denoted by the point 'a' called out on the plot. The expected uncompensated error rate is 0.59%. While it may be intuitive to decrease the modulation order to improve the error performance, the results in Fig. 2 show this is not advisable; a BPSK link under the same conditions is operating beyond the dispersion limit, and has a high error rate of about 7.5%, even in the effective total absence of noise. Rather, if the modulation order is increased to 16, then Fig. 3 shows the error rate is decreased to 0.035% for the same SNR of 20 dB, more than an order of magnitude improvement (denoted by the point 'b' called out on the plot). In fact, the results presented in Fig. 3 show that when the link distance is above 8.5 km (shown by the vertical dashed line) and the SNR greater than 20 dB, switching from 4-QAM to 16-QAM will always improve the error rate, due to the greater spectral efficiency (and thus lower bandwidth and GVD) of the 16-QAM scheme.

For a second illustration, now consider the 16-QAM link operating at 14 km in Fig. 4, with a high SNR of 40 dB. The expected error rate is 0.148% (denoted by point 'a' on the plot). To improve this error rate, the default choice might be to decrease the modulation order, but again, this worsens the error rate to a value of 9.5% for 4-QAM

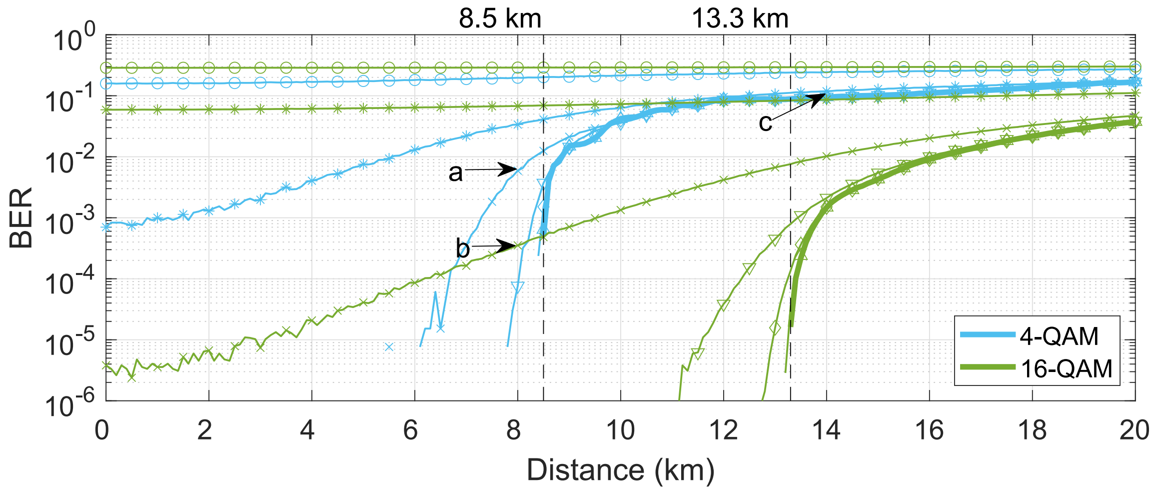


Fig. 3 – Reverse waterfall plot for *M*-QAM links, *M* = 4, 16, over distances from 0 to 20 km with various SNRs. SNR values of 0, 10, 20, 30, 40, and 50 dB are designated by markers ○, ⁎, ×, ▽, ◇, and △ respectively, and infinite SNR is denoted by a thick, solid line with no marker. Atmospheric conditions are the same as in Fig. 2. The fact that the 40 dB, 50 dB, and infinite SNR curves are almost indistinguishable because beyond 40 dB, dispersion is the dominant source of errors, rather than noise.

(according to point 'c' on Fig. 3). However, increasing the modulation order to 64 offers improvement according to Fig. 4, reducing the error rate to 4.7×10^{-5} (point 'b' on the plot). In this case, the improvement again relies on the increased spectral efficiency of the higher order modulation scheme, with the stipulation that the SNR be 40 dB or greater. This exacting constraint on SNR arises from the small spacing between symbol decision boundaries for the higher-order link, made even stricter by the fact that dispersion, though reduced, has still shifted some of the received symbols closer to the decision boundaries.

Thirdly, notice from Fig. 2 that, for the 60 Gb/s case presented here, there are some modulation schemes that, in general, constitute poor choices for a link without dispersion compensation. Namely, a 256-QAM scheme has a worse error rate at all distances and SNRs than either 16-QAM or 64-QAM (with the exception of a slight and insignificant region around 14 km in the noise-free case, where it has performance marginally better than the 16-QAM scheme). While BPSK and 4-QAM do underperform 256-QAM over long distances with higher SNR, there are also two other modulation types (16-QAM and 64-QAM) that outperform 256-QAM in nearly all situations, so while 256-QAM is an improvement over some modulation schemes, it is never the best choice (and this holds true for all higher SNRs as well).

At this point in the discussion, there are a few assumptions that need to be addressed. One point of concern may be that in most applications, the bandwidth of a wireless link is fixed and the data rate varies with modulation order, while in the results we present, the *band-width* varies with modulation order while data rate is held constant. However, we are not proposing this is how future terahertz links should operate; indeed, variable bandwidth channels would be both an engineering and a regulatory challenge, and probably are not appropriate for most circumstances. Rather, we chose to present our data in this manner because allowing bandwidth to vary with modulation order makes the counter-intuitive behavior of the spectrally-efficient, low bandwidth links (that is, 256-QAM) the most clear and explicit. This does not change our simulation results; it is just a data- presentation choice. Varying the bandwidth of the link was the best way to show that decreasing the bandwidth does not necessarily decrease the error rate, and that modulation type becomes an important factor due to atmospheric GVD.

Finally, all discussion up until this point has been focused on communication links in which dispersion is uncompen- sated. Though GVD has not historically been a concern for wireless communication systems (owing to the com- paratively narrow bandwidth of legacy microwave com- munication links), it has been extensively investigated in fiber optics, where dispersion-compensating technology is relatively mature. Additionally, there are other forms of temporal dispersion that *have* been identified, studied, and compensated in existing wireless links, which often arise from multipath propagation. Although atmospheric GVD is a new phenomena for wireless links, dispersion in general is not. This may lead some to think that since dispersion can and has been compensated by both photonic [27] and electronic [28, 29] means, then these technologies would be readily adapted for use in a terahertz wireless communication system. Specifically, it might be assumed digital signal processing filters, also known as equalizers, will be able to compensate GVD and thus render the problem of GVD irrelevant.

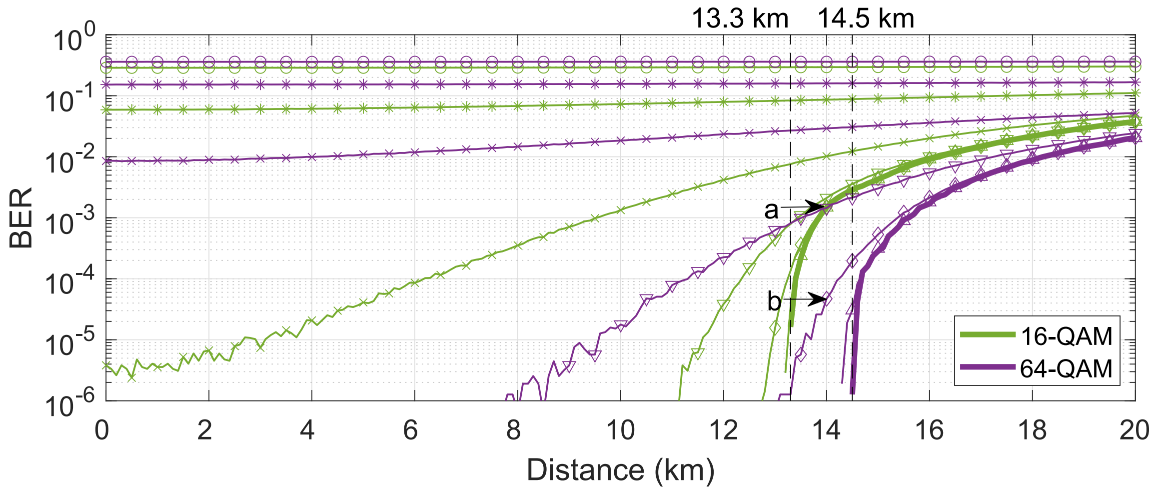


Fig. 4 – Reverse waterfall plot for M-QAM links, M = 16, 64, over distances from 0 to 20 km with various SNRs. SNR values of 0, 10, 20, 30, 40, and 50 dB are designated by markers ○, ⁎, ×, ▽, ◇, and △ respectively, and infinite SNR is denoted by a thick, solid line with no marker. Atmospheric conditions are the same as in Fig. 2.

While equalizers are certainly *theoretically* capable of compensating dispersion, whether they will be physically realizable for terahertz frequencies (and, if so, when) is still yet to be determined. There are still questions that remain to be answered before we can confidently assert which equalizer architectures will be most suited to operation in the terahertz bands. In 4G architectures utilizing orthogonal frequency division multiplexing, equalizers operate on channels at most 20 MHz wide, and this is the dominant wireless technology. However, in the terahertz bands, the signal bandwidth may be up to 100 GHz, potentially over three orders of magnitude larger! Even the fiber optic equalizers referenced previously typically have bandwidths less than 100 GHz [30]. This high bandwidth significantly complicates filter design. If terahertz sub-bands are kept only a few tens of kilohertz wide in order to avoid this problem, then the number of sub-bands (and thus equalizers) scales up by potentially four orders of magnitude. Further complications include 10 to 100 times greater Doppler shifts, noise bandwidths two to three orders of magnitude larger, and dispersion profiles that change with weather, not to mention the is- sues of receiver linearity, phase noise, and dynamic range which are already challenges for 3G and 4G hardware [31]. While none of this changes the fact that dispersion is theoretically reversible, it does raise the question: are current equalization algorithms and the digital hardware on which they are implemented capable of performing the task? Current research is presently being undertaken to investigate these issues [32], and bottlenecks related to sampling rate and signal processing limitations have been identified [33, 34]. Presently, it seems premature to assume that the equalization and signal processing technologies we currently have will carry over to terahertz channels without significant modification and innovation. Accordingly, we have not included equalization routines in our simulations for concern they would produce results that are not necessarily realistic. Furthermore, we wish to limit the scope of this paper to a characterization and description of the GVD induced by the atmosphere, and its interaction with modulation type. If and when dispersion compensating technology is implemented in future terahertz communication systems, the judicious selection of modulation type will reduce the performance requirements placed on such technology by utilizing modulation schemes naturally resistant to dispersion-induced bit errors. This could be especially important for relaxing the signal processing burden in terahertz transceivers.

4. CONCLUSION

In this work, we leveraged highly accurate models of the atmosphere to predict the effects of uncompensated atmospheric GVD on the bit error rate of high-capacity terahertz links using various orders of M-QAM. A significant finding was that, due to GVD, unintuitive situations arise in which higher-order modulations offer superior error rate performance than lower-order modulations. This is contrary to what would be expected in a traditional wireless link with a lower bandwidth, in which the selection of modulation type is dominated by the SNR alone. It is anticipated that this will need to be taken into account by both future link designers and adaptive modulation algorithms attempting to select the ideal modulation scheme for present channel conditions. A related finding was that, in uncompensated links, there are some modulation orders that should not be used (or are at least never the best choice). Specifically, high-order modulations, such as 256-QAM (and above) suffer from stringent requirements on both SNR and maximum allowable symbol shift due to dispersion, which when combined lead to suboptimal performance for all or nearly all combinations of links distance and SNR.

5. ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. 1746055 and by the National Science Foundation under Grant No. 2018110. Any opinions, findings, and conclusions or recommendations ex- pressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] Walid Saad, Mehdi Bennis, and Mingzhe Chen. "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems". In: *IEEE network* 34.3 (2019), pp. 134–142.

[2] GMDT Forecast. "Cisco visual networking index: global mobile data traffic forecast update, 2017–2022". In: *Update* 2017 (2019), p. 2022.

[3] L. Chettri and R. Bera. "A Comprehensive Survey on Internet of Things (IoT) Toward 5G Wireless Systems". In: *IEEE Internet of Things Journal* 7.1 (2020), pp. 16–32. DOI: 10 . 1109 / JIOT . 2019 . 2948888.

[4] K. Shafique, B. A. Khawaja, F. Sabir, S. Qazi, and M. Mustaqim. "Internet of Things (IoT) for Next-Generation Smart Systems: A Review of Current Challenges, Future Trends and Prospects for Emerging 5G-IoT Scenarios". In: *IEEE Access* 8 (2020), pp. 23022–23040. DOI: 10.1109/ACCESS.2020.2970118.

[5] Y. Xing and T. S. Rappaport. "Propagation Measurement System and Approach at 140 GHz-Moving to 6G and Above 100 GHz". In: *2018 IEEE Global Communications Conference (GLOBECOM)*. 2018, pp. 1–6. DOI: 10.1109/GLOCOM.2018.8647921.

[6] Ho-Jin Song and Tadao Nagatsuma. "Present and future of terahertz communications". In: *IEEE transactions on terahertz science and technology* 1.1 (2011), pp. 256–263.

[7] A Hirata, T Kosugi, H Takahashi, J Takeuchi, K Murata, N Kukutsu, Y Kado, S Okabe, T Ikeda, F Suginosita, et al. "5.8-km 10-Gbps data transmission over a 120-GHz-band wireless link". In: *2010 IEEE International Conference on Wireless Information Technology and Systems*. IEEE. 2010, pp. 1–4.

[8] Hiroyuki Takahashi, Toshihiko Kosugi, Akihiko Hirata, Jun Takeuchi, Koichi Murata, and Naoya Kukutsu. "120-GHz-band fully integrated wireless link using QSPK for realtime 10-Gbit/s transmission". In: *IEEE transactions on microwave theory and techniques* 61.12 (2013), pp. 4745–4753.

[9] Cheng Wang, Changxing Lin, Qi Chen, Bin Lu, Xianjin Deng, and Jian Zhang. "A 10-Gbit/s wireless communication link using 16-QAM modulation in 140-GHz band". In: *IEEE Transactions on Microwave Theory and Techniques* 61.7 (2013), pp. 2737–2746.

[10] Xinying Li, Jianjun Yu, Kaihui Wang, Yuming Xu, Long Chen, Li Zhao, and Wen Zhou. "Bidirectional delivery of 54-Gbps 8QAM W-band signal and 32-Gbps 16QAM K-band signal over 20-km SMF-28 and 2500-m wireless distance". In: *Optical Fiber Communication Conference*. Optical Society of America. 2017, Th5A–7.

[11] Ingmar Kallfass, Jochen Antes, Axel Tessmann, Thomas Zwick, and Ralf Henneberger. "Multi-Gigabit high-range fixed wireless links at high millimeterwave carrier frequencies". In: *2017 IEEE Radio and Wireless Symposium (RWS)*. IEEE. 2017, pp. 45–48.

[12] Qiuyu Wu, Changxing Lin, Bin Lu, Li Miao, Xin Hao, Zhaohui Wang, Yi Jiang, Wenqiang Lei, Xianjing Den, Hongbin Chen, et al. "A 21 km 5 Gbps real time wireless communication system at 0.14 THz". In: *2017 42nd International Conference on In- frared, Millimeter, and Terahertz Waves (IRMMW- THz)*. IEEE. 2017, pp. 1–2.

[13] M. Polese, J. M. Jornet, T. Melodia, and M. Zorzi. "Toward End-to-End, Full-Stack 6G Terahertz Networks". In: *IEEE Communications Magazine* 58.11 (2020), pp. 48–54. DOI: 10 . 1109 / MCOM . 001 . 2000224.

[14] S. Ghafoor, N. Boujnah, M. H. Rehmani, and A. Davy. "MAC Protocols for Terahertz Communication: A Comprehensive Survey". In: *IEEE Communications Surveys Tutorials* 22.4 (2020), pp. 2236–2282. DOI: 10.1109/COMST.2020.3017393.

[15] T. Schneider, A. Wiatrek, S. Preussler, M. Grigat, and R. Braun. "Link Budget Analysis for Terahertz Fixed Wireless Links". In: *IEEE Transactions on Terahertz Science and Technology* 2.2 (2012), pp. 250–256. DOI: 10.1109/TTHZ.2011.2182118.

[16] K. Rikkinen, P. Kyosti, M. E. Leinonen, M. Berg, and A. Parssinen. "THz Radio Communication: Link Budget Analysis toward 6G". In: *IEEE Communica- tions Magazine* 58.11 (2020), pp. 22–27. DOI: 10 . 1109/MCOM.001.2000310.

[17] A. A. Boulogeorgos, E. N. Papasotiriou, and A. Alexiou. "A Distance and Bandwidth Dependent Adaptive Modulation Scheme for THz Communications". In: *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. 2018, pp. 1–5. DOI: 10.1109/SPAWC.2018.8445864.

[18] C. Han and I. F. Akyildiz. "Distance-Aware Bandwidth-Adaptive Resource Allocation for Wireless Systems in the Terahertz Band". In: *IEEE Transactions on Terahertz Science and Technology* 6.4 (2016), pp. 541–553. DOI: 10 . 1109 / TTHZ . 2016.2569460.

[19] C. Han and I. F. Akyildiz. "Distance-aware multi-carrier (DAMC) modulation in Terahertz Band communication". In: *2014 IEEE International Confe- rence on Communications (ICC)*. 2014, pp. 5461–5467. DOI: 10.1109/ICC.2014.6884190.

[20] Pei Yang, Zeliang Ou, and Hongwen Yang. "Capacity of AWGN and Rayleigh Fading Channels with *M*-ary Inputs". In: *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*. IEEE. 2018, pp. 1–6.

[21] Shi Jia, Xianbin Yu, Hao Hu, Jinlong Yu, Pengyu Guan, Francesco Da Ros, Michael Galili, Toshio Morioka, and Leif K. Oxenløwe. "THz photonic wireless links with 16-QAM modulation in the 375-450 GHz band". In: *Opt. Express* 24.21 (Oct. 2016), pp. 23777–23783. DOI: 10.1364/OE.24.023777. URL: http : / / www . opticsexpress . org /abstract.cfm?URI=oe-24-21-23777.

[22] Chong Han and Yi Chen. "Propagation Modeling for Wireless Communications in the Terahertz Band". In: *IEEE Communications Magazine* 56.6 (2018), pp. 96–101. DOI: 10.1109/MCOM.2018.1700898.

[23] Yihong Yang, Mahboubeh Mandehgar, and Daniel R. Grischkowsky. "Understanding THz Pulse Propagation in the Atmosphere". In: *IEEE Transactions on Terahertz Science and Technology* 2.4 (2012), pp. 406–415. DOI: 10.1109/TTHZ.2012.2203429.

[24] John F O'Hara and Daniel R Grischkowsky. "Comment on the veracity of the ITU-R recommendation for atmospheric attenuation at terahertz frequencies". In: *IEEE Transactions on Terahertz Science and Technology* 8.3 (2018), pp. 372–375.

[25] Yihong Yang, Mahboubeh Mandehgar, and D Grischkowsky. "Determination of the water vapor continuum absorption by THz-TDS and Molecular Response Theory". In: *Optics Express* 22.4 (2014), pp. 4388–4403.

[26] Karl Strecker, Sabit Ekin, and John OHara. *Fundamental Performance Limits on Terahertz Wireless Links Imposed by Group Velocity Dispersion*. 2021. arXiv: 2104.00611 [eess.SP].

[27] Karl Strecker, Sabit Ekin, and John F O'Hara. "Compensating atmospheric channel dispersion for terahertz wireless communication". In: *Scienti· ic reports* 10.1 (2020), pp. 1–8. DOI: 10.1038/s41598–020–62692–7.

[28] Julián S Bruno, Vicenç Almenar, Javier Valls, and Juan L Corral. "Real-time 20.37 Gb/s optical OFDM receiver for PON IM/DD systems". In: *Optics express* 26.15 (2018), pp. 18817–18831.

[29] Amit Raikar, Amol Jirage, and Ashwini Narake. "A SURVEY: DISPERSION COMPENSATION TECHNIQUES FOR OPTICAL FIBER COMMUNICATION". In: *International Journal* 15 (2019), 16th.

[30] Xing Ouyang, Giuseppe Tall, Mark Power, and Paul Townsend. "Experimental demonstration of 112 Gbit/s orthogonal chirp-division multiplexing based on digital up-conversion for IM/DD systems with improved resilience to system impairments". In: *2018 European Conference on Optical Communication (ECOC)*. IEEE. 2018, pp. 1–3.

[31] Somayeh Mohammady, Ronan Farrell, David Malone, and John Dooley. "Performance investigation of peak shrinking and interpolating the PAPR reduction technique for LTE-advance and 5G signals". In: *Information* 11.1 (2020), p. 20.

[32] Priyangshu Sen, Dimitris A. Pados, Stella N. Batalama, Erik Einarsson, Jonathan P. Bird, and Josep M. Jornet. "The TeraNova platform: An integrated testbed for ultra-broadband wireless communications at true Terahertz frequencies". In: *Computer Networks* 179 (2020), p. 107370. ISSN: 1389-1286. DOI: https : / / doi . org / 10 . 1016 / j . comnet . 2020 . 107370. URL: https : / / www . sciencedirect . com / science / article / pii /S1389128620304473.

[33] Viduneth Ariyarathna, Arjuna Madanayake, and Josep Miquel Jornet. "Real-Time Digital Baseband System for Ultra-Broadband THz Communication". In: *2020 45th International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz)*. 2020, pp. 1–2. DOI: 10 . 1109 / IRMMW - THz46771 . 2020.9370733.

[34] Hadi Sarieddeen, Mohamed-Slim Alouini, and Tareq Y Al-Naffouri. "An overview of signal processing techniques for terahertz communications". In: *arXiv preprint arXiv:2005.13176* (2020).

AUTHORS

Karl L. Strecker received his B.S. in 2018 and M.S. in 2020, both in electrical engineering from Oklahoma State University. He is currently pursuing his Ph.D. degree in electrical engineering at the same institution. From 2018 to the present, he has worked as a research assistant in the Ultrafast Terahertz and Optoelectronics Laboratory at Oklahoma State University. His research interests include wireless communications, group velocity dispersion management, and terahertz material characterization. Mr. Strecker was a recipient of the 2020 National Science Foundation Graduate Research Fellowship.

Sabit Ekin received a B.Sc. degree in electrical and electronics engineering from Eskişehir Osmangazi University, Turkey, in 2006, an M.Sc. degree in electrical engineering from New Mexico Tech, Socorro, NM, USA, in 2008, and a Ph.D. degree in electrical and computer engineering from Texas A&M University, College Station, TX, USA, in 2012. He was a visiting research assistant with the Electrical and Computer Engineering Program, Texas A&M University at Qatar from 2008 to 2009. In summer 2012, he was with the Femtocell Interference Management Team in the Corporate Research and Development, New Jersey Research Center, Qualcomm Inc. He joined the School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK, USA, as an assistant professor, in 2016. He has four years of industrial experience from Qualcomm Inc., as a Senior Modem Systems Engineer with the Department of Qualcomm Mobile Computing. At Qualcomm Inc., he has received numerous Qualstar awards for his achievements/contributions on cellular modem receiver design. His research interests include the design and performance analysis of wireless communications systems in both theoretical and practical point of views, interference modeling, management and optimization in 5G, mmWave, HetNets, cognitive radio systems and applications, satellite communications, visible light sensing, communications and applications, RF channel modeling, non-contact health monitoring, and Internet of Things applications.

John F. O'Hara (M'05 – SM'19) received his BSEE degree from the University of Michigan in 1998 and his Ph.D. (electrical en-gineering) from Oklahoma State University (OSU) in 2003. Hewas a Director of Central Intel-ligence Postdoctoral Fellow at Los Alamos National Laboratory(LANL) until 2006. From 2006-2011 he was with the Center for Integrated Nanotechnologies (LANL) and worked on numerous metamaterial projects involving dynamic control over chirality, resonance frequency, polarization, and modulation of terahertz waves. In 2011, he founded a consulting/research company, Wavetech, LLC specializing in automation and IoT devices. In 2017 he joined OSU as an assistant professor in the School of Electrical & Computer Engineering. His current research involves terahertz wireless communications, terahertz sensing and imaging with metamaterials, IoT, and light-based sensing and communications. He has 3 patents and around 100 publications in journals and conference proceedings. Dr. O'Hara is a Senior Member of IEEE.

TERAHERTZ WIRELESS COMMUNICATIONS IN SPACE

Meltem Civas[1] and Ozgur B. Akan[1,2]

[1]Next-generation and Wireless Communications Laboratory (NWCL), Department of Electrical and Electronics Engineering, Koç University, Istanbul, Turkey , [2]Internet of Everthing (IoE) Group, Department of Engineering, University of Cambridge, UK

NOTE: Corresponding author: Meltem Civas, mcivas16@ku.edu.tr

Abstract – *The New Space Era has increased communication traffic in space by new space missions led by public space agencies and private companies. Mars colonization is also targeted by crewed missions in the near future. Due to increasing space traffic near Earth and Mars, the bandwidth is getting congested. Moreover, the downlink performance of the current missions is not satisfactory in terms of delay and data rate. Therefore, to meet the increasing demand in space links, Terahertz band (0.1-10 THz) wireless communications are proposed in this study. In line with this, we discuss the major challenges that the realization of THz band space links pose and possible solutions. Moreover, we simulate Mars-space THz links for the case of a clear Mars atmosphere, and a heavy dust storm to show that even in the worst conditions, a large bandwidth is available for Mars communication traffic.*

Keywords – Deep space communications, inter-satellite links, terahertz communications

1. INTRODUCTION

The start of a *New Space Era* has led to a paradigm shift in the space industry. An increasing number of private com- panies with space missions have emerged, and spacecraft are getting smaller and easily deployable with the help of enabling technologies. Moreover, the non-terrestrial networks are also recognized by the future release 17 of the 3rd Generation Partnership Project (3GPP) to be studied under the study item ”Non-Terrestrial Networks”. Vertical networks are expected to be integrated into the next generation 5G and beyond networks. This can lead to many novel paradigms including the Internet of Space Things (IoST) [1]. Therefore, available bandwidths are getting congested. National Aeronautics and Space Ad- ministration (NASA) estimates that the growth factor of deep space communication capability should be at least 10 to meet the growing demand in the next three decades [2]. Therefore, high data rate communication technolo- gies are required. Free space optical (FSO) communica- tion has been envisioned for space applications. However, optical communication highly depends on atmospheric conditions such as fog, cloud, and haze where optical links experience strong attenuation [3]. Moreover, FSO is costly, and beam alignment problems pose several addi- tional challenges. Terahertz band (0.1-10 THz) wireless communications, which can enable high data rates on the order of Terabits per second, is an alternative to FSO.

The advantages of FSO such as large bandwidth, high data rate, high security, no spectrum licensing are of growing interest. NASA’s demonstration in 2013 called Lunar Laser Communication Demonstration proved that high downlink and uplink transmission rates from lunar orbit to Earth are possible via optical links and NASA will further conduct Deep Space Optical Communications demonstration as a part of Psyche mission to show that higher data rates, 10-100 times of the current state of the art, are attainable via optical wireless communi- cations [4]. The European Data Relay Satellite System (EDRS) also employs optical inter-satellite links, which can reach a data rate of 1.8 Gbps for a transmission distance up to 45000 km [5]. Apart from institutional projects, the companies such as SpaceX and Google aim to use opti- cal communications for space-to-space or air-to-air links of their projects under development. Despite its advan- tages, FSO communications have still several challenges. Regarding space links, due to high transmission distances Effective Isotropic Radiated Power (EIRP) must be high. This can be enabled by large aperture transmitting op- tics, which result in narrow-beam divergence ($\propto \lambda/D$, where λ is the wavelength and D is the diameter of aper- ture). Thus, a deviation of the optical beam from the tar- get results in an outage. Considering the same transmit- ting antenna size, THz systems have looser beam pointing requirements compared to FSO links, since we consider larger wavelengths.

FSO communications are also highly affected by atmo- spheric conditions such as clouds, haze, fog, and atmo- spheric turbulence. For instance, when visibility is less than 50 m when dense fog exists, attenuation can be as high as 350 dB/km, which limits transmission [3]. In case of fog with the same visibility, there are several win- dows in the THz band where the attenuation is below 100 dB/km [6]. Compared to FSO, THz communications are more robust to weather effects in general. Although the attenuation in THz links for long-distance communication is still a challenge to be

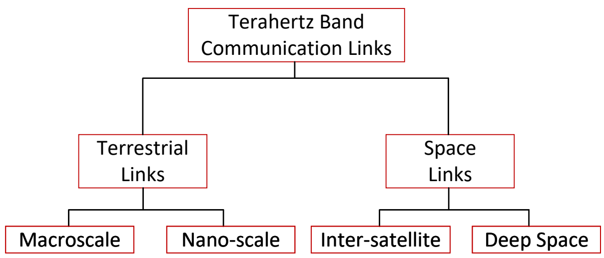


Fig. 1 – Classification of THz links [8].

addressed, there are promising solutions such as large THz antenna arrays enabling ex- tremely high gains. Moreover, THz waves are not affected by turbulence-induced scintillation, which is observed as intensity fluctuations at the signal, as much as FSO links [7, 3].

THz band communication links can be classified into two as terrestrial and non-terrestrial networks as shown in Fig. 1. Terrestrial networks comprise macroscale and nano-scale links. In this study, we consider non-terrestrial network components ground/space and space links consisting of inter-satellite and deep space communication links. We aim to identify the challenges related to THz space links and discuss the possible solutions. The realization of THz space links poses several challenges. The spreading loss due to the expansion of propagating electromagnetic waves is increasing drastically with the frequency. This limits the communication distance to few meters on Earth due to immature THz source technology, which can enable transmit power on the order of milliwatts. Moreover, strong molecular absorption results in high atmospheric attenuation in Earth-to-space links; thus, limits the utilization of the high THz band. Artificial satellites occupy higher atmospheric layers of Earth or deep space where the air molecules are scarce or none. Therefore, THz inter-satellite links do not experience significant atmospheric attenuation. Regarding another terrestrial planet Mars, atmospheric attenuation is expected to be low compared to Earth because water molecules, which are the primary source of atmospheric attenuation, are scarce in the Mars atmosphere. These create an opportunity of utilizing the high THz band, consequently providing high data rates. In line with this, later we simu- late the transmittance of Mars's atmosphere in clear and dusty atmospheric conditions using an accurate radiative transfer tool called Planetary Spectrum Generator (PSG) to show the availability of a large bandwidth for Mars communication.

The rest of the paper is organized as follows. In Section 2, we describe the applications of THz space links. In Section 3, we discuss the challenges THz band communications encounter, and then in Section 4, we simulate zenith transmittance of Mars atmosphere. In Section 5, conclusions are stated.

2. APPLICATIONS

Utilizing THz space links can pave the way for novel applications, some of which are discussed as follows.

2.1 Earth observation

Earth observation using artificial satellites began with the launch of Sputnik 1 by the former Soviet Union in 1957. Since then, many Earth observation satellites have been launched. Most of these satellites occupy Low Earth Orbits (LEO) and transmit a large amount of data to Earth daily. Recently, an increasing number of LEO satellites are being launched so that the bandwidth used is get ting congested. To reduce the transmission delays and support the transmission of a large amount of sensing data to Earth, technologies supporting high data rates are required. FSO communications, providing connectivity within a few kilometers using laser beams, have been proposed as a viable solution [9]. For instance, EDRS employs FSO communications between LEO satellites collecting Earth observation data and GEO satellite relaying data to Earth. However, with the start of a New Space Era, satellites are getting miniaturized, and deploying many small satellites, e.g., CubeSats, is preferred [10]. The power and size requirements of FSO systems far exceed the limitations of cube/micro/nano-satellites. On the other hand, building THz transceivers with a large number of antenna arrays, i.e., phased Multiple-Input and Multiple-Output (MIMO) arrays, in a small footprint, is possible thanks to novel materials such as graphene [11]. Thus, for long-distance and high-data-rate near-Earth transmission, THz communications can be leveraged in the future. Large THz arrays are also advantages over FSO links in terms of beam-alignment, e.g., they can provide automatic alignment by their scanning ability [12]. However, there exist issues to realize such THz transceivers. The main impediments include the lack of practical THz signal sources and detectors, implementation and optimization of antenna arrays [12].

2.2 Interplanetary communications by hybrid THz/FSO links

Current state-of-the-art communication technologies used are not able to support high data rate interplanetary communications as a part of space information networks, which can result in numerous applications including space observation, Internet of Things (IoT), and maritime monitoring. To illustrate, Mars Reconnaissance Orbiter employs X-band (8-12 GHz) and Ka-band (26.5-40 GHz) to communicate with the NASA Deep Space Network, which comprises deep space communication facilities for commanding and tracking spacecraft. The data rate is between 0.5 and 4 megabits per second [13]. The services such as live video feeding, high-resolution scientific data streaming, virtual reality for controlling rovers and other machines, and real-time data transmission will require much higher data rates. Although latency

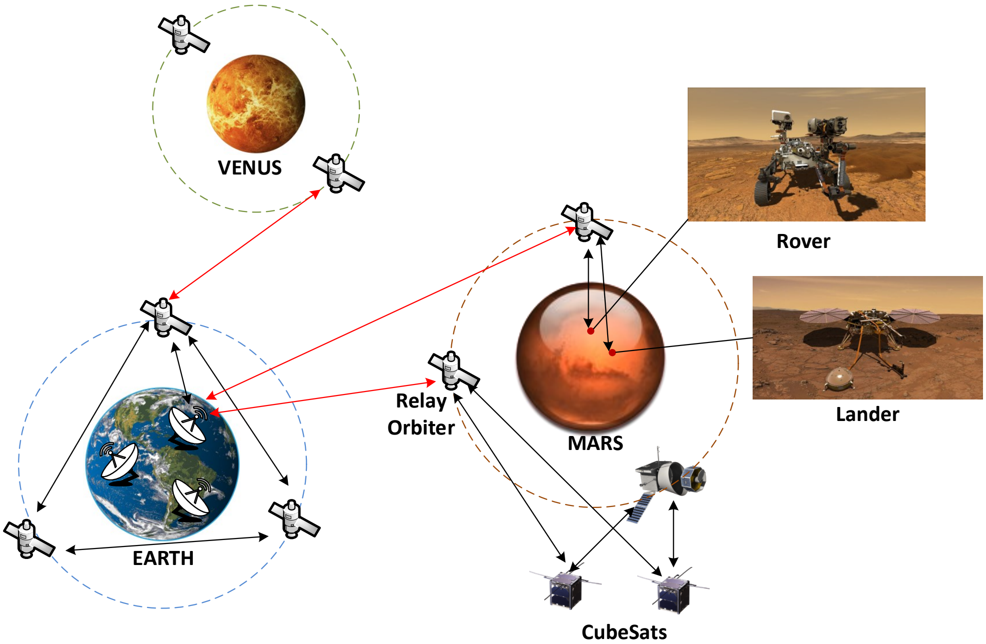


Fig. 2 – Interplanetary communications: Black arrows (THz links), red arrows (FSO links).

is still an issue, these services can be enabled by hybrid THz/FSO links as a part of space information networks. In case of an outage due to beam pointing errors or strong atmospheric attenuation, THz links can be used as a backup [14] since THz links are more robust to weather conditions and pointing errors.

THz communication capability can be integrated into FSO communication architecture to support high data rates and can lead to novel applications such as space exploration. For instance, Mars rovers communicate with Earth through relay orbiters since the availability of orbiters is much longer than that of rovers. A Mars rover can communicate with a relay orbiter through THz links and the relay orbiter relays messages using FSO links because THz communications are more robust compared to FSO in ground/space links. CubeSats are being used in interplanetary missions during mission-critical events, which are exemplified by Mars Cube One (MarCO). CubeSats can be equipped with THz transceivers in the future so that they can communicate with the landing spacecraft using low latency THz links and, to relay information to Earth using FSO links.

3. CHALLENGES AND SOLUTIONS

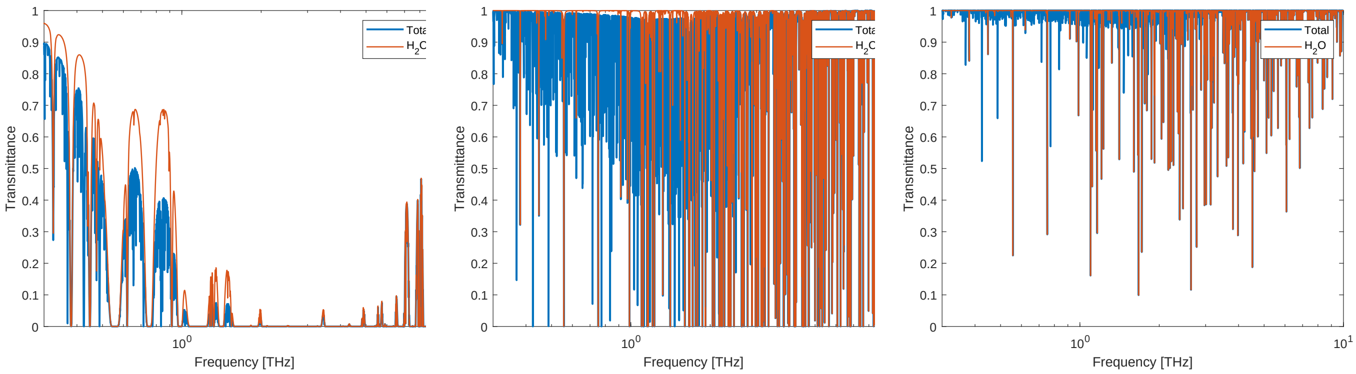
In this section, we discuss the several challenges THz space links encounter. These challenges include molecular absorption and spreading loss, interference to passive services. We also discuss the potential directions to address these challenges.

3.1 Molecular absorption loss

Molecular absorption loss, which occurs when the part of wave energy is transformed into molecular energy due to the vibration of molecules, is one of the main impediments affecting Earth-space THz links. Water vapor molecules, which are scarce in the atmosphere of Mars, are the primary sources of molecular absorption on Earth in THz frequencies [15]. Inter-satellite THz links among Low Earth Orbits (LEO), Medium Earth Orbits (MEO), and Geosynchronous Orbit (GEO) satellites are not affected by molecular absorption loss because they operate at the altitudes where water vapor is almost none. To combat the high atmospheric attenuation on Earth, several approaches have been proposed. In the following section we give an overview of these approaches.

3.1.1 Dry ground sites

Atmospheric Precipitable Water Vapor (PWV) is the primary cause of strong atmospheric attenuation at THz frequencies. Considering ground-based telescopes of THz radio astronomy such as Atacama Large Millimeter/Submillimeter Array (ALMA) located at a high and dry plateau of Chile and Combined Array for Research in Millimeter-Wave Astronomy (CARMA) which was operating in the United States, Suen *et al.* identified the loca- tions on Earth with lowest water vapor [15]. Dry sites for radio astronomy as well as satellite communications in- clude Antarctica, Greenland, the Atacama Desert, and the Tibetan Plateau. Numerous dry sites that can provide acceptable performance have been identified in the United States and Europe as well.



(a) Altitude = 0 km. **(b)** Altitude =16km. **(c)** Altitude =35km.

Fig. 3 – The transmittance of Earth-space links at several altitudes from one of the driest locations of the Earth, i.e., ALMA, denoting the fraction of EM radiation after experiencing molecular absorption. The zenith angle is 35°.

Suen et al. further investigated the performance of a THz ground to geostationary satellite links [16]. They have shown that utilizing radio astronomy platforms with large aperture antenna arrays, which are located at dry sites of Earth, for satellite communications 1 terabit/second link performance can be exceeded in clear atmosphere conditions [16]. In [17], the authors show that 1 terabit/second is attainable in the low THz band for ground/satellite links utilizing massive antenna arrays and establishing the ground stations at Tanggula, Tibet, where PWV is very low.

3.1.2 *High and low altitude platforms*

To combat high atmospheric attenuation in ground/satellite THz links, placing transceivers on airborne platforms has been proposed in [16, 18]. High and low altitude platforms such as aerostats, aircraft, and high altitude balloons can be employed for transceiver placement. These platforms need a lower aperture diameter compared to ground-based transceivers and can also offer performances comparable to ground-based platforms with large apertures since they operate at the altitudes where water vapor density is low [16]. In Fig. 3, Earth-space link transmittances for various altitude levels in a dry location of Chile, which are simulated in the Planetary Spectrum Generator (PSG) [19], is depicted. According to this, numerous bands, which are not feasible to use at sea level, are available for use at high altitudes.

3.1.3 *Hybrid ground/satellite links*

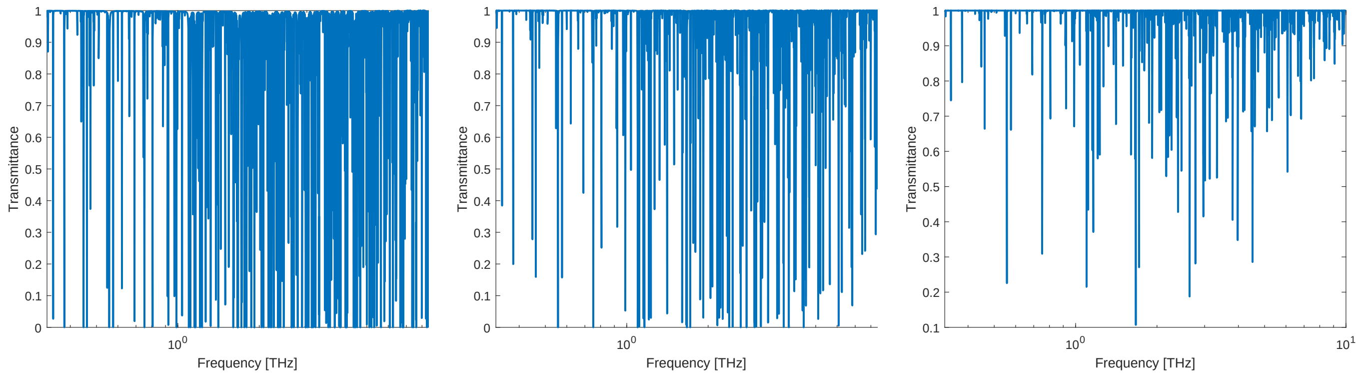
Placing ground stations only at dry locations can limit the potential of THz communication. Akyildiz *et al.* propose ground-satellite links enabled by microwave (e.g., X band (8-12 GHz), Ku band (12-18 GHz, Ka-band (26.5- 40 GHz)), and mm-Wave/THz bands for small satellites called CubeSats in [20]. If a ground/satellite link is not suitable for transmission at THz frequencies, mmWave and microwave bands can be utilized. The idea is based on sending a pilot signal to obtain the availability of the link according to some criteria (e.g., Line of Sight (LOS) and weather conditions). The analysis in [20] shows that even with a relatively high water vapor density, data rates on the order of tens of Gbps can be achieved.

3.2 Spreading loss

As an Electromagnetic (EM) wave propagates through a medium it expands and this leads to a loss called spreading loss. Spreading loss is one of the significant challenges limiting communication to short distances at THz frequencies because free space path loss increases with the frequency in a quadratic relation according to Friis' law. Regarding space links, we consider the distances at least on the order of thousands of kilometers. On the other hand, transmit power in THz frequencies is on the order of milliwatts due to immature THz source technology, which is also called the THz gap. Thus, high gain antennas with high directivity are required for THz space links. Several approaches for combating the problem of high propagation loss have been proposed in the literature [21]. Some of these solutions apply to indoor and nano-scale communications such as intelligent surfaces controlling the behavior of an EM wave [22] and graphene plasmonic nano-antennas [23]. In the following section, we discuss the potential solutions for THz space links.

3.2.1 *Radio astronomy optics*

Large aperture THz optics, which are exemplified by ALMA comprising 54 reflector antennas with 12 meter diameter and 12 smaller antennas with a 7 meter dish diameter, are already being used by radio astronomy. In line with this, one approach is employing large aperture THz ground stations and airborne stations with smaller apertures [16]. Large apertures can provide high gain; however, one drawback is that the construction cost increases with the diameter [18].



(a) Altitude = 0 km. **(b)** Altitude =16km. **(c)** Altitude =30km.

Fig. 4 – Zenith transmittance of clear Mars atmosphere for various altitudes.

3.2.2 *Reflect-arrays*

Reflect-array antennas are being used by the systems such as satellite communications, radars and deep-space communication links [24]. Traditional aperture antennas can provide high gain but they are not as electronically flexible as phased arrays. However, the implementation cost of phased arrays is high. Reflect-arrays offering high gain, low cost, ease of manufacturing as well as electronic flexibility are a compromise between aperture antennas and phased arrays. To illustrate, the MarCO spacecraft communicated with Earth at a distance of 160 million km on X-band via a high-gain reflect-array antenna with small volume [25]. Operating frequencies of reflect-array antennas are now shifting towards THz frequencies [26]. Thus, they offer the potential to be employed in THz space links. On the other hand, enabling technologies need to be studied because RF and MEMS technologies such as semiconductor diodes and MEMS lumped elements do not apply to the THz band due to loss and size constraints [27].

3.2.3 *Ultra-massive MIMO*

The concept of Ultra-Massive Multiple Input Multiple Output (UM-MIMO) has been introduced in [28] for combating the distance problem in THz communications. According to this concept, using novel materials such as graphene to build antennas with a number of antenna elements in a small footprint is possible. Utilizing both space and frequency, the coverage range can be increased. However, the realization of this concept poses several challenges [28]. The performance of UM-MIMO depends on the THz channel; thus, accurate THz channel mod- els are required. Moreover, to control arrays, dynamic beam-forming algorithms are needed [20]. Moreover, the performance of THz MIMO links can also be affected by the frequency-dependent diffraction of THz waves, which arises from the divergence of THz beams from their modulation side-bands. This results in degraded bit-error- rate performance due to the detection of unwanted spec- trum information. Thus, novel detection and demodula- tion methods are required in this direction [29].

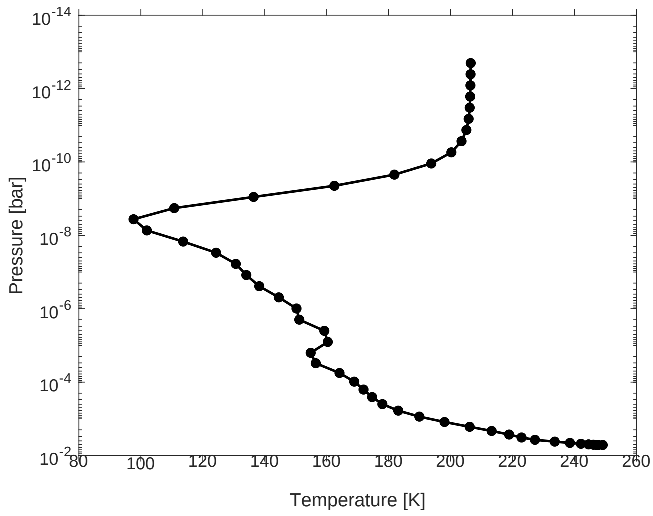
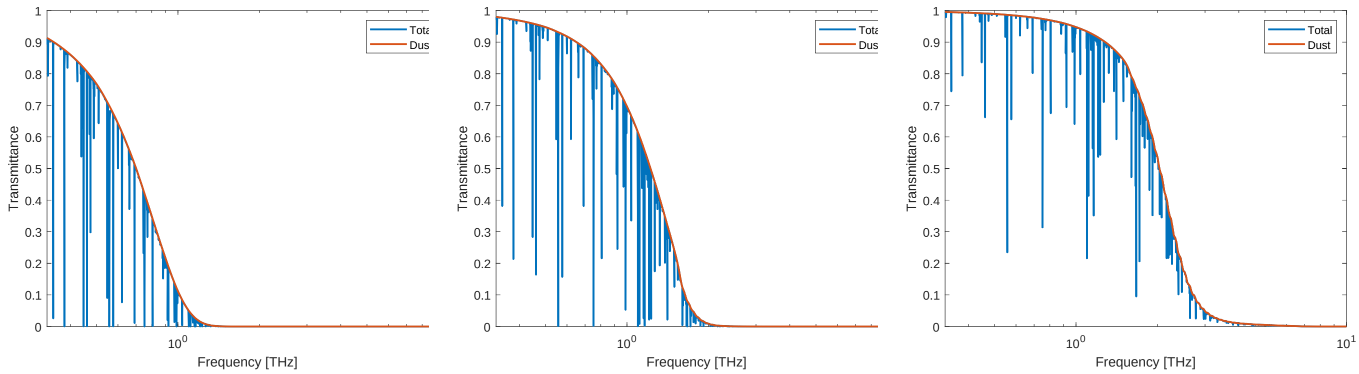


Fig. 5 – Pressure vs temperature.

3.3 Spectrum sharing

Spectrum beyond 275 GHz is not largely regulated in the Radio Regulations (RR). Footnote 5.565 of RR identifies numerous frequency bands in the range from 275 GHz to 1000 GHz that are used by passive services, namely Radio Astronomy Services (RAS), Earth Exploration Satellite Services (EESS), and Space Research Service, and states that the activity of these services must be protected from harmful interference of active services until the frequency allocation is established [30]. Active and passive services will coexist on the spectrum. Current spectrum sharing studies, which aim to identify the bands where the coexistence of active and passive services is possible, mainly focus on the interference to EESS [31] because RAS telescopes are located in high dry mountains. However, coexistence studies should cover both RAS and EESS because both can be affected by THz ground/satellite and inter-satellite links.



(a) Altitude $= 0$ km. **(b)** Altitude $=16$km. **(c)** Altitude $=30$km.

Fig. 6 – Zenith transmittance of Mars atmosphere during a local dust storm for various altitudes ($r_{eff} = 1.5\mu m$).

4. TERAHERTZ MARS-SPACE LINKS

Preparing Mars for human exploration is one of the targets of current missions on Mars. Thus, communication among human explorers, remote-controlled vehicles, space instruments, or any other space entities will be an essential part of Mars missions soon. Water vapor molecules and oxygen are scarce in the Mars atmosphere. Thus, the effect of molecular absorption is less compared to Earth. However, Mars's atmosphere can pose other challenges due to scattering aerosols such as seasonal dust storms, limiting reliable communication.

Table 1 – Vertical profile of Mars atmosphere

Gas	Symbol	Composition
Carbon Dioxide	CO_2	95.717%
Nitrogen	N_2	1.991%
Oxygen	O_2	0.152%
Carbon Monoxide	CO	818.452 ppm
Water Vapor	H_2O	194.232 ppm
Ozone	O_3	4.750 ppb
Column	-	1.947e+27 m-2
Col mass	-	1.408e+2 kg/m2

4.1 The Planetary Spectrum Generator

The PSG is an online radiative-transfer suite, which can generate planetary spectra of planets and other plane- tary objects [19]. PSG uses several radiative transfers and scattering models, and databases including spectroscopic (e.g., high-resolution transmission database (HITRAN)) and climatological databases (e.g., Mars Climate Database (MCD)). We use the module of PSG extracting atmospheric profile MCD. The model of the Mars atmosphere comprises 49 layers up to 257.90 km. PSG further extracts in- formation of pressure and temperature, and profiles of at- mospheric gases, scattering particle sizes from MCD [19]. The pressure and temperature profile used is shown in Fig. 5. The surface temperature is 279.48 K and the sur- face pressure is 5.1666 mbar. Regarding the geometry, we have used the looking-up mode, in which the zenith path is considered while integrating radiative transfer. We consider the location with the longitude of 175.5 and the latitude of -14.8 (Mars Exploration Rover A landing site). The zenith angle is 25.045 and the azimuth angle is 297.909 in the simulations. Mars date considered is 2018/05/07.

4.2 Zenith transmittance

For a clear Mars atmosphere, the zenith transmittance of a Mars-space link is shown in Fig. 4. When the altitude is 0 km, it can be observed that in the $0.3 - 10$ THz band, transmittance values are greater than 0.9. There are numerous sharp decreases in the band ($1 - 10$ THz). At the altitude of 30 km, the whole band is available except for sharp decreases in the transmittance, since molecular absorption is not so effective at high altitudes due to the very low abundance of molecules.

Dust storms are an important phenomenon of Mars. They can be classified as local ($< \ 2000\text{km}^2$), regional ($> \ 2000\text{km}^2$) and planet-encircling [32]. Global dust storms occur in the southern spring and summer seasons of Mars. Dust devils are common phenomena seen both on Earth and Mars. They inject a high amount of dust into the atmosphere. During Mars southern spring and summer 0.9 to 2.9×10^{11} kg/km^2 dust devil flux is estimated in [33]. The amount of dust injected into the atmosphere in local and regional dust storms is also reported com- parable to dust devils. Accordingly, in a local storm the abundance of dust in the atmosphere can be calculated as $0.145 \times 10^6\text{kg}/\text{m}^2$ (for 2000km^2 area) for the worst case. Mean or effective radius r_{eff} is another important parameter to examine the effect of dust on the scattering. We assume the abundance of dust is $0.145 \times 10^6\text{kg}/\text{m}^2$ in every 49 layers of the atmosphere. According to observation of Mars atmosphere, r_{eff} varies between $1.5\mu m$ and $1.6\mu m$ [34]. Thus, we consider the effect of a local dust storm on THz-band transmission in a Mars-space link at various altitudes for $r_{eff} = 1.6\mu\text{m}$. The results in Fig. 6

show that even in a heavy dust storm, transmittance values are close to 0.9 in the low THz band when the altitude is 0, which is the worst case. When the altitude is increased, transmit- tance values are higher than 0.9 up to 1 THz.

5. CONCLUSION

THz space links can pave the way for services including live video feeding, high-resolution imagery, and virtual reality from space. Apart from its inherent challenges such as molecular absorption loss, THz band communications pose other challenges due to developing THz source and antenna technologies. Therefore, in this study, we discuss the major challenges of THz space links, namely molecular absorption loss, spreading loss, and interference from RAS and EESS. The possible solutions for ground/space links are locating ground stations to high and dry locations, and multiband antennas. Regarding spreading loss, large aperture antennas can be used for ground stations benefiting from radio astronomy optics. Reflect-arrays and UM-MIMO are other solutions.

ACKNOWLEDGEMENT

This work was supported in part by Huawei Graduate Research Scholarship.

REFERENCES

[1] Ian F Akyildiz and Ahan Kak. "The Internet of Space Things/CubeSats". In: *IEEE Network* 33.5 (2019), pp. 212–218.

[2] NASA. *Deep Space Communications*. URL: `https://scienceandtechnology.jpl.nasa.gov/research/research-topics-list/communications-computing-software/deep-space-communications`.

[3] Hemani Kaushal and Georges Kaddoum. "Optical communication in space: challenges and mitigation techniques". In: *IEEE communications surveys & tutorials* 19.1 (2016), pp. 57–96.

[4] NASA. *Deep Space Optical Communications (DSOC)*. URL:`https://www.nasa.gov/mission_pages/tdm/dsoc`.

[5] ESA. *European data relay satellite system (edrs) overview*. URL:`https://artes.esa.int/edrs/overview`.

[6] Thomas Schneider, Andrzej Wiatrek, Stefan Preußler, Michael Grigat, and Ralf-Peter Braun. "Link budget analysis for terahertz fixed wireless links". In: *IEEE Transactions on Terahertz Science and Technology* 2.2 (2012), pp. 250–256.

[7] Chong Han, Yongzhi Wu, Zhi Chen, and Xudong Wang. "Terahertz Communications (TeraCom): Challenges and Impact on 6G Wireless Systems". In: *arXiv preprint arXiv:1912.06040* (2019).

[8] Meltem Civas, Turker Yilmaz, and Ozgur B Akan. "Terahertz Band Intersatellite Communication Links". In: *Next Generation Wireless Terahertz Communication Networks*. CRC Press, pp. 337–354.

[9] Vineeth Venugopal. "The light way to go 6G". In: *Photonics Focus* 2.1 (2021), pp. 30–33.

[10] Oltjon Kodheli, Eva Lagunas, Nicola Maturo, Shree Krishna Sharma, Bhavani Shankar, JF Montoya, JC Duncan, Danilo Spano, Symeon Chatzinotas, Steven Kisseleff, et al. "Satellite Communications in the New Space Era: A Survey and Future Challenges". In: *arXiv preprint arXiv:2002.08811* (2020).

[11] Meltem Civas, Oktay Cetinkaya, Murat Kuscu, and Ozgur B Akan. "Universal Transceivers: Opportunities and Future Directions for the Internet of Everything (IoE)". In: *arXiv preprint arXiv:2107.01028* (2021).

[12] Kaushik Sengupta, Tadao Nagatsuma, and Daniel M Mittleman. "Terahertz integrated electronic and hybrid electronic–photonic systems". In: *Nature Electronics* 1.12 (2018), pp. 622–635.

[13] NASA. *Communications with Earth*. URL: `https://mars.nasa.gov/mro/mission/communications/`.

[14] Mohammad Ali Khalighi and Murat Uysal. "Survey on free space optical communication: A communication theory perspective". In: *IEEE communications surveys & tutorials* 16.4 (2014), pp. 2231–2258.

[15] Jonathan Y Suen, Michael T Fang, and Philip M Lubin. "Global distribution of water vapor and cloud cover sites for high-performance THz applications". In: *IEEE Transactions on Terahertz Science and Technology* 4.1 (2014), pp. 86–100.

[16] Jonathan Y Suen, Michael T Fang, Sean P Denny, and Philip M Lubin. "Modeling of terabit geostationary terahertz satellite links from globally dry locations". In: *IEEE Transactions on Terahertz Sci ence and Technology* 5.2 (2015), pp. 299–313.

[17] Ruchen Zhen and Chong Han. "Link Budget Analysis for Massive-Antenna-Array-Enabled Terahertz Satellite Communications". In: *Journal of Shanghai Jiaotong University (Science)* 23.1 (2018), pp. 20–27.

[18] Jonathan Y Suen. "Terabit-per-second satellite links: a path toward ubiquitous terahertz communication". In: *Journal of Infrared, Millimeter, and Terahertz Waves* 37.7 (2016), pp. 615–639.

[19] Geronimo L Villanueva, Michael D Smith, Silvia Protopapa, Sara Faggi, and Avi M Mandell. "Planetary Spectrum Generator: An accurate online radiative transfer suite for atmospheres, comets, small bodies and exoplanets". In: *Journal of Quantitative Spectroscopy and Radiative Transfer* 217 (2018), pp. 86–104.

[20] Ian F Akyildiz, Josep M Jornet, and Shuai Nie. "A new CubeSat design with reconfigurable multi-band radios for dynamic spectrum satellite communication networks". In: *Ad Hoc Networks* 86 (2019), pp. 166–178.

[21] Ian F Akyildiz, Chong Han, and Shuai Nie. "Combating the distance problem in the millimeter wave and terahertz frequency bands". In: *IEEE Communications Magazine* 56.6 (2018), pp. 102–108.

[22] Shuai Nie, Josep M Jornet, and Ian F Akyildiz. "Intelligent environments based on ultra-massive MIMO platforms for wireless communication in millimeter wave and terahertz bands". In: *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2019, pp. 7849–7853.

[23] Michele Tamagnone, JS Gomez-Diaz, Juan Ramon Mosig, and Julien Perruisseau-Carrier. "Reconfigurable terahertz plasmonic antenna concept using a graphene stack". In: *Applied Physics Letters* 101.21 (2012), p. 214102.

[24] Sean Victor Hum and Julien Perruisseau-Carrier. "Reconfigurable reflectarrays and array lenses for dynamic antenna beam control: A review". In: *IEEE Transactions on Antennas and Propagation* 62.1 (2013), pp. 183–198.

[25] Richard E Hodges, Nacer Chahat, Daniel J Hoppe, and Joseph D Vacchione. "A Deployable High-Gain Antenna Bound for Mars: Developing a new folded-panel reflectarray for the first CubeSat mission to Mars." In: *IEEE Antennas and Propagation Magazine* 59.2 (2017), pp. 39–49.

[26] Eduardo Carrasco and Julien Perruisseau-Carrier. "Reflectarray antenna at terahertz using graphene". In: *IEEE Antennas and Wireless Propagation Letters* 12 (2013), pp. 253–256.

[27] Diego Correas-Serrano and J Sebastian Gomez-Diaz. "Graphene-based antennas for terahertz systems: A review". In: *arXiv preprint arXiv:1704.00371* (2017).

[28] Ian F Akyildiz and Josep Miquel Jornet. "Realizing ultra-massive MIMO (1024× 1024) communication in the (0.06–10) terahertz band". In: *Nano Communication Networks* 8 (2016), pp. 46–54.

[29] Jianjun Ma, Nicholas J Karl, Sara Bretin, Guillaume Ducournau, and Daniel M Mittleman. "Frequency-division multiplexer and demultiplexer for terahertz wireless links". In: *Nature communications* 8.1 (2017), pp. 1–8.

[30] Radio Regulations. "Articles". In: Edition of (2016).

[31] Thomas Kü rner. "Regulatory Aspects of THz Communications and related Activities towards WRC 2019". In.

[32] Christian Ho, Nasser Golshan, and Arvydas Kliore. "Radio wave propagation handbook for communication on and around Mars". In: (2002).

[33] Patrick L Whelley and Ronald Greeley. "The distribution of dust devil activity on Mars". In: Journal of Geophysical Research: Planets 113.E7 (2008).

[34] Michael D Smith. "Spacecraft observations of the Martian atmosphere". In: Annu. Rev. Earth Planet. Sci. 36 (2008), pp. 191–219.

AUTHORS

Meltem Civas received her B.S. degree in electrical and electronics engineering in Bilkent University, Ankara, Turkey, in January 2016, and her M.S. degree at the Electrical and Electronics Engineering Department, Koç University, Istanbul, Turkey in September 2018. She is currently a research assistant at the Next-generation and Wireless Communications Laboratory and pursuing her Ph.D. degree in Electrical and Electronics Engineering, Koç University. Her research interests include Internet of Everything and molecular communications.

Ozgur B. Akan received his Ph.D. degree in electrical and computer engineering from the Broadband and Wireless Networking Laboratory, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta in 2004. He is currently with the Electrical Engineering Division, Department of Engineering, University of Cambridge, United Kingdom, and also the director of the Next-Generation and Wireless Communications Laboratory in the Department of Electrical and Electronics Engineering, Koç University. His research interests include wireless, nano and molecular communications, and the Internet of Everything.

FUNDAMENTAL LIMITS OF HIGH-EFFICIENCY SILICON AND COMPOUND SEMICONDUCTOR POWER AMPLIFIERS IN 100-300 GHz BANDS

James F. Buckwalter[1], Mark J. W. Rodwell[1], Kang Ning[1], Ahmed Ahmed[1], Andrea Arias-Purdue[1], Jeff Chien[1], Everett O'Malley[1] and Eythan Lam[1]

[1]University of California – Santa Barbara, 1160 Harold Frank Hall, Santa Barbara, CA 93106

NOTE: Corresponding author: James Buckwalter, buckwalter@ucsb.edu

Abstract *– This paper reviews the requirements for future digital arrays in terms of power amplifier requirements for output power and efficiency and the device technologies that will realize future energy-efficient communication and sensing electronics for the upper millimeter-wave bands (100-300 GHz). Fundamental device technologies are reviewed to compare the needs for compound semiconductors and silicon processes. Power amplifier circuit design above 100 GHz is reviewed based on load line and matching element losses. We present recently presented class-A and class-B PAs based on a InP HBT process that have demonstrated record efficiency and power around 140 GHz while discussing circuit techniques that can be applied in a variety of integrated circuits.*

Keywords – Digital array, high-efficiency, millimeter-wave, power amplifier

1. INTRODUCTION

Frequencies between 100-300 GHz, known as Upper millimeter-Wave (UmmW) bands, offer an opportunity for convergence of communication and sensing systems to support future high-throughput backhaul and radar applications [1]. In particular, frequency bands located at 140 and 220 GHz feature O2 and H2O absorption windows for low propagation loss in outdoor channel environments [2]. Digital array applications in UmmW bands require mature electronic and packaging technologies and previously Integrated Circuits (IC) demonstrated poor power efficiency and higher package costs when compared to lower millimeter-wave (LmmW) bands (28/39/60 GHz). While other bands, including the 60 GHz bands offer substantial bandwidth, the high absorption at 60 GHz prohibits energy efficient operation over more than a kilometer and UmmW offers opportunity for high-bandwidth.

Moreover, the UmmW bands offer shorter wavelength relative to LmmW and this feature allows more Transmit (TX) and Receive (RX) elements within a given aperture area. The array spacing at 140 GHz would be approximately 1 mm and, therefore, a 1cm x 1cm array could host around 100 elements while a 28 GHz array might have only 4 elements in the same aperture area. Consequently, the UmmW beam-former array will contain a relatively large number of steerable elements that might be packed into the small form factor, controlled with independent digitally-controlled Baseband (BB) and Intermediate Frequency (IF), and this poses a large-scale integration challenge that must be solved with unified design that includes IC, packaging, and device technologies.

The large number of array elements in the UmmW array suggests design architecture based on digital array techniques rather than traditional RF beam-forming approaches that leverage signal processing techniques based on massive MIMO (mMIMO) for higher spatial resolution than conventional MIMO systems [3]. Reusing time-frequency resources across multiple users can ultimately support higher spectral efficiency across a network and with the available bandwidth in UmmW devices link capacity might approach 1 Terabit/second [4]. Moreover, a large number of antennas will focus energy into small regions in the space. Thus, in theory, the transmit power can be reduced while maintaining a high Signal-to-Noise Ratio (SNR), resulting in higher spatial energy efficiency. Of course, there is a circuit overhead to generate the RF signals across the mMIMO array which scales linearly with the number of elements. At some point, a larger array incurs substantial power consumption penalties. Early demonstrations of mMIMO in sub-6 GHz based on commercially-available software-defined radios require kilowatts in signal processing [5].

Early work on line-of-sight MIMO in millimeter-wave bands was demonstrated nearly a decade ago [4]. As millimeter-wave systems are reaching commercial adoption based on RF beam-formers, interest has pivoted into the further capabilities of digital beam-formers in conventional 5G bands for mMIMO and preliminary studies have begun to demonstrate multi-beam test beds [4]. However, the feasibility of mMIMO systems above 100 GHz requires evaluation of the benefit from both signal processing and the short wavelength of the signals, particularly for systems where area is a limitation.

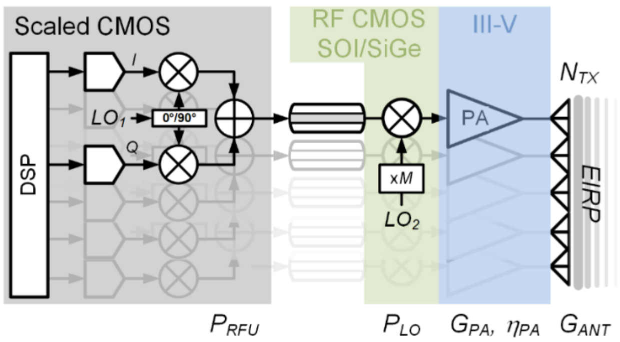


Fig. 1 - Architectural partition for large-scale integration of transmitters in UMMW bands where 2D arrays prohibit direct integration of the full transceiver behind a single antenna element.

The architecture of a UmmW digital beam-former is illustrated in Fig. 1 to move the DSP and Intermediate Frequency (IF) signal generation blocks in scaled CMOS away from the 2D transmit (or receive) front end in an array. An RF upconverter allows the IF (or baseband) to be shifted to the RF band where a Power Amplifier (PA) generates the appropriate output power level. This architecture might be substantially different than a sub-6 GHz or 28 GHz array where the element spacing is relaxed relative to the amount of integration on a single silicon chip and multiple polarizations and even transmit/receive might be fit within a single array site.

This paper reviews the Transmitter (TX) and PA constraints for digital beam-formers that would support future mMIMO deployments and details the output power requirements and how these can be translated into an efficient PA design using commercially available III-V and Si transistor technologies today [6]. Recent work supports the possible rapid improvement in circuit design above 100 GHz that may usher in an era of energy efficient communication and sensing electronics for the UmmW bands. Section 2 will review the digital beam-former architecture and highlight the different demands in the Base Station (BS) and the User Equipment (UE). Section 3 reviews the fundamental limits on efficiency for device technologies above 100 GHz and highlights the transistor improvements. Section 4 discusses advanced III-V transistor nodes and the possibilities in these nodes. Section 5 discusses how attempts to seek optimal transistor embedding networks to improve gain struggle to improve efficiency, and Section 6 presents recent work that has demonstrated how one particular process, InP HBT, has the potential to significantly increase efficiency in the 100-300 GHz bands.

2. DIGITAL BEAM-FORMING

In the UmmW bands, the BS and UE for a communication link would have different requirements for the PA. To produce multiple beams for different users, the digital beam-former in the BS is illustrated in Fig. 1. As opposed to an RF beam-former, each PA is supported by an individual pair of DACs and RF Upconverter (RFU) unit. The architecture translates into higher Peak to-Average Power Ratio (PAPR) requirements in the digital beam-former for the PA. By simulating the sum of random uncorrelated QPSK waveforms associated with several users, the aggregate PAPR asymptotically approaches 12 dB. High PAPR suggests relatively high peak power and linearity requirements for the PA.

The UmmW band poses a unique problem for the packaging of a digital transmit array. To occupy a 2D array site of only 1mm by 1mm, very compact electronic ICs must fit into a 2-Dimensional (2D) array without a 3D packaging solution [7]. More importantly, the power consumption of a UmmW front end is constrained to prevent a substantial thermal load that must be dissipated with the small area with the large numbers of elements.

The PA poses the most significant problem for the energy efficiency of a UmmW digital beam-forming array. While an all-silicon solution might be attractive to integrate the large number of elements, the PA performance in silicon processes is limited in terms of output power and efficiency. Prior work demonstrated single digit efficiency in CMOS and SiGe processes for 15-20 dBm output power in the power amplifier [8][9][10]. With single digit efficiency, the power requirements across 100s of elements are tremendous, particularly given the dense array of power amplifiers at 140 GHz, and lends to an insurmountable heat removal challenge. Consequently, the transmit digital beam-former illustrated in Fig. 1 might include a III-V PA if the power and efficiency of another device technology supports the requirements.

A notional link between a BS and UE can be mathematically analyzed to understand the digital beam-former power consumption. As more channels are added, the digital beam-former increases the RFU and DAC power linearly. However, the PA output power decreases given a fixed EIRP constraint.

By adding the power consumption of the RFU (which includes DAC and IF signal generation) and the LO power (which includes LO distribution, multiplier, buffer and mixer power consumption), an optimization is found based on basic circuit parameters for the PA (drain) efficiency and the overhead of the per-element RF upconversion.

$$N_{OPT} = \sqrt{\frac{\eta_{MOD} EIRP}{\eta_{PA} G_{ANT}(P_{RFU}+P_{LO})}} \quad (1)$$

Minimum energy is found when the PA power consumption offsets the overhead power consumption for the RFU and DAC. Not only is the power consumption significantly reduced, the size of the array can be reduced by a factor of two through the smaller number of elements. When we consider this optimization in terms of the number of elements, the optimum output power per element is

$$P_{OUT,OPT} = \frac{\eta_{PA}}{\eta_{MOD}}(P_{RFU} + P_{LO}) \quad (2)$$

Consequently, the optimum output power is a fraction of the DC power consumed in the RFU and LO based on the PA efficiency.

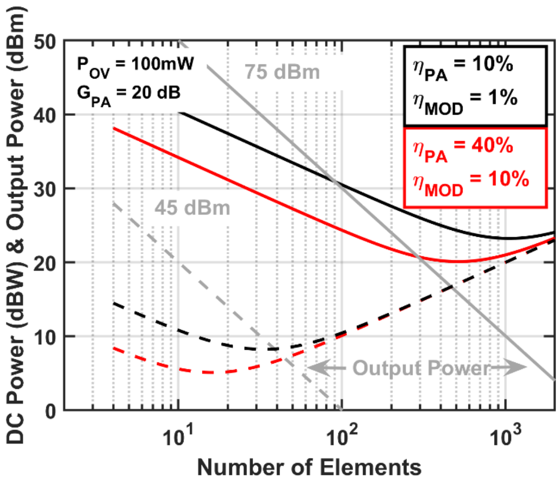


Fig. 2 - Optimization of the number of elements under different EIRP conditions demonstrating the output power per element in dBm and the total array power consumption in dBW. (Assumes G_{ANT} = 5 dB).

As an example, the minimum energy of the TX array is plotted in terms of the number of elements in Fig. 2. A total overhead power consisting of the RFU and LO is assumed to be 100 mW. Additionally, we assume the PA gain is 20 dB. Using current state-of-the-art numbers for the DC to RF efficiency of the modulator and the PA, 1% and 10%, respectively, the base station (EIRP = 75 dBm) is optimized for more than 1000 elements. Moreover, the minimum power is around 220 W. As expected from (2), the output power per PA is around 10 dBm based on the overhead of 100 mW and the efficiency of 10%. On the other hand, if the PA efficiency is improved to 40%, the number of elements reduces (by roughly a factor of 2) but the optimum PA output power increases to around 16 dBm.

On the other hand, the UE (EIRP = 45 dBm) is minimized for 32 elements at a total power of 7 W and output power per element again of 10 dBm. This is the same as the BS since the assumptions in (2) do not change between the BS and UE in this exercise. With improvements in the PA efficiency, the number of elements decreases to 16 elements and the overall power is around 3 W and the average power per element is again 16 dBm. The critical insight is that improvements in efficiency also increase the average output power demands per element.

While these power consumptions for the array might seem high, considering that the BS MIMO beam-former could support more than 100 users (load factor of ¼), the power is amortized by the throughput to each user. If each user receives a 10Gb/s QPSK stream, the overall energy efficiency of each BS beam amounts to around 200 pJ/b which is comparable to the energy efficiency of communications at lower frequency bands. Notably, the UE achieves similar energy efficiency.

Consequently, the PA requirements in the BS or UE are not excessively different nor are they particularly onerous from the standpoint of design compared to LmmW. Most interestingly, the design demands higher output power per element as the efficiency improves and suggests examination of the device technologies for the TX PA. For the purposes of further analysis, we will consider a peak power of 20 dBm as a target per element output power to account for the PAPR.

3. POWER AMPLIFIER CONSTRAINTS

The Power Added Efficiency (PAE) and output power are directly related to properties of the underlying device technology. The PAE can be expressed in terms of several factors.

$$\mathrm{PAE} = \eta\left(1-\frac{1}{G}\right)\left(1-\frac{V_K}{V_{DD}}\right)\left(\frac{Q_o}{Q_o+Q_t}\right) \qquad (3)$$

The PAE depends on the drain efficiency, η, the operating gain of the PA (G), the knee (V_K) and supply voltage (V_{DD}) of the device, and the loss factor for matching the load line of the device to the load impedance, which is expressed above in terms of a impedance transformation quality factor, (Q_t), and passive element quality factor (Q_o). Consequently, the knee voltage relative to the supply voltage imposes a penalty on the available PAE. Additionally, the impedance transformation between the load line of the transistor and the output impedance, e.g. 50 Ohms, reduces the maximum efficiency.

The drain efficiency is determined by the biasing of the PA as well as the harmonic tuning at the load. With only matching of the load line of the device to the load and no additional voltage waveform shaping at the PA output, the gate bias determines the drain efficiency as a class of operation. The conduction angle, θ, of the drain current captures the maximum drain efficiency. When the device is conducting during the entire period (θ = 2π), the transistor is operating in class A with maximum drain efficiency of 50%. If the bias is reduced such that the transistor conducts half the time (θ = π/2), the drain current is class B and the maximum drain efficiency increases to 78%. Unfortunately, the reduced conduction in class B also reduces that transistor gain. Conduction angles between class A and B are referred to as AB.

The available power gain produces a limitation in PAE at bands near the maximum cutoff frequency of the transistor, fmax. For PAs operating above 100 GHz, fmax is often not much larger than the frequency of operation based on currently available device technologies. The available gain is, therefore, limited and the class of operation can be chosen as part of the optimization process.

Fig. 3 indicates the theoretical PAE as a function of the conduction angle. Note that several parameters in (3) are functions of θ, including the shape factor, but also the gain and impedance matching. As the θ reduces from class A to class B and beyond into class C, the PAE increases and then collapses as the gain drops. Notably, several factors in the PAE change as a function of the θ. The load-line impedance increases the loss of the matching network. Moreover, as one moves from LmmW bands at 60 GHz to the UmmW bands at 140 and 220 GHz, the optimum conduction angle moves away from class B bias towards the class A bias. The maximum possible PAE drops from more than 40% to 30% at 140 GHz. When the PA design targets 220 GHz, the maximum PAE becomes around 17%.

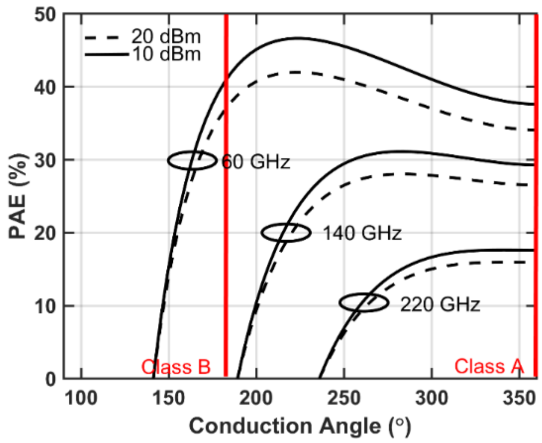


Fig. 3 - PAE as a function of power amplifier conduction angle for upper millimeter-wave frequencies. (f_{max}/f_T = 400 GHz, V_K = 0.7, V_{DD} = 2.5, Q = 10).

We can also compare output power requirements in the previous section for 20 dBm and 10 dBm output power. At 60 GHz, the lower power PA is capable of 7% better efficiency. However, once we reach 220 GHz, the benefit of the reduced output power is smaller. The difference in the PAE achievable with different power levels is attributed to the change in the impedance matching networks and additional losses. Consequently, the dominant performance limitation on PAE for UmmW PAs is the available gain to realize high efficiency at the moderate output powers described in Section 2. We will investigate approaches to improve the gain while optimizing the PAE factors in (3) in the next two sections.

4. UMMW SEMICONDUCTOR TECHNOLOGY COMPARISON

We can study approximate parameters of available processes to understand the PAE limit in (3) with different trade-offs in terms of available gain, voltage handling requirements, and load-line matching conditions for a given matching or output power condition.

We assume that the passive elements have similar quality factor. Table 1 illustrates sample characteristics of different transistor technologies that are available for operation above 100 GHz in III-V and SiGe/SOI CMOS technologies. Si CMOS technologies at 22-nm and 28-nm might also be considered for the benefit of digital integration but typically do not offer substantially different performance in UmmW bands than 65-nm or CMOS SOI processes and the process parameters are generic. The roadmap of RF-optimized InP HBT processes has been discussed in [11] and [12]. Current 250-nm InP HBT processes are capable of fmax exceeding 600 GHz while being relatively mature with commercial applications. Scaling to 130-nm and beyond can yield fmax exceeding 1 THz. The evolution of SiGe HBTs has also produced remarkable fmax increases that have reached similar speeds to InP [13]. SiGe BiCMOS processes have been optimized for digital and RF performance [14]. Current processes offer several differentiated HBTs in a single process optimized for breakdown and fT/fmax. CMOS SOI processes based on partially depleted SOI substrates have evolved from a digital process to RF-optimized approaches with high-resistivity substrates and RF back end-of-the-line [15]. CMOS and CMOS SOI offer similar supply and knee voltage. The 40-nm GaN HEMT process is described in [16] and offers a 400-GHz fmax. The characteristics of the different processes is summarized in Table 1 with an emphasis on commercially available processes with the highest f_{max}.

Table 1 - Comparison of UMMW semiconductor device technologies

Technology	f_{max} (GHz)	V_{SUP} (V)	V_K (V)	I_{MAX} (A/mm)	P_{RF} (W/mm)	R_{LL} (Ωmm)
InP HBT	600	2.5	0.7	3	1.4	1.2
SiGe HBT	450	1.3	0.5	2.2	0.44	0.7
CMOS	310	1.1	0.3	1	0.2	1.6
GaN HEMT	400	12	2	1.6	4.0	12.5

Based on the supply voltage and the knee voltage, the maximum output power can be calculated from $P_{RF} = \frac{1}{4}(V_{DD} - V_K)I_{MAX}$ while the load-line resistance is $R_{LL} = 2(V_{DD} - V_K)/I_{MAX}$. The RF output power and load-line resistance is shown in Table 1 when normalized to 1 mm. The high supply voltage of the GaN HEMT due to breakdown characteristics suggests high power density and load-line resistance relative to the other processes.

We compare these technologies in two different ways to understand device selection for high efficiency. On one hand, we would choose a device that offers a load line close to 50 Ohm to avoid loss in the matching network. On the other hand, we consider a power target and investigate the required size of the transistor. A larger transistor introduces design challenges to distribute the RF power into and out of the transistor. A large device, relative to the wavelength, typically incurs a drop in the potential fmax. At 140 GHz, a device width/length of more than 200 um would pose significant conditions to distribute the signal.

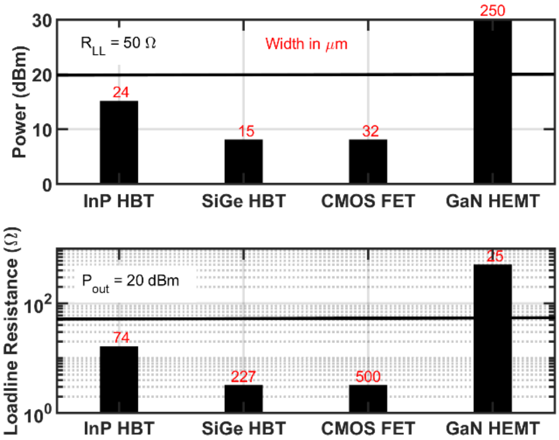


Fig. 4 - Comparison of process technology trade-offs under fixed resistance (50 Ohm) and fixed power (20 dBm) conditions.

First, to minimize the loss of the impedance matching to the load line, we might choose the device geometry (i.e. width) to provide a 50 Ohm load line. The top plot in Fig. 4 indicates the output power that is developed by each device technology. For instance, the InP process will produce 15 dBm while the SiGe process will produce roughly 6 dBm. The GaN HEMT would deliver around 30 dBm output power. For a 20-dBm target power outlined in Section 2, the InP HBT and GaN HEMT are closest to the target for a 50-Ohm match.

Table 2 - Theoretical efficiency bounds for 20 dBm output power at 140 GHz

Technology	InP HBT	SiGe HBT	CMOS FET	GaN HEMT
PAE (Q = 1000)	45%	34%	32%	43%
PAE (Q = 10)	39%	24%	23%	34%
Conduction Angle	203°	222°	260°	232°

Second, we might also compare the technologies for a fixed output power such as 20 dBm in Fig. 4. The device presenting a load-line matching condition closest to 50 Ohms is the most desirable from the standpoint of PAE. Notably, the InP HBT is the best choice as GaN HEMT presents a large load-line resistance while a CMOS FET is a very low load-line matching.

Based on these process parameters and the analysis of dependence on conduction angle, a preliminary estimate of the PAE can be gathered for different technologies in Table 2 along with the optimal conduction angle. These values were calculated based on equation (3) and searching for the maximum PAE versus conduction angle as shown from Fig. 3. Since both the shape factor, gain, and impedance transformation depend on conduction angle, we must consider all these factors to understand the class of operation that will achieve the highest PAE. The PAE is computed for a passive quality factor of 10 and 1000. Notably, InP can theoretically reach more than 40% efficiency with a deep class AB/B bias while GaN might approach similar efficiency. Silicon processes should be able to exceed 30%.

5. THEORETICAL COMPARISONS AGAINST PUBLISHED WORK

To compare the insights into the device performance bounds on published PAs above 100 GHz, we surveyed PA results from all published work including CMOS, SOI CMOS, SiGe HBTS, InP HBTs, GaN HEMTs, and GaAs mHEMTs during the previous two decades to establish trends and future development possibilities for more efficient radio and millimeter-wave systems in the UmmW (100-300 GHz) band.

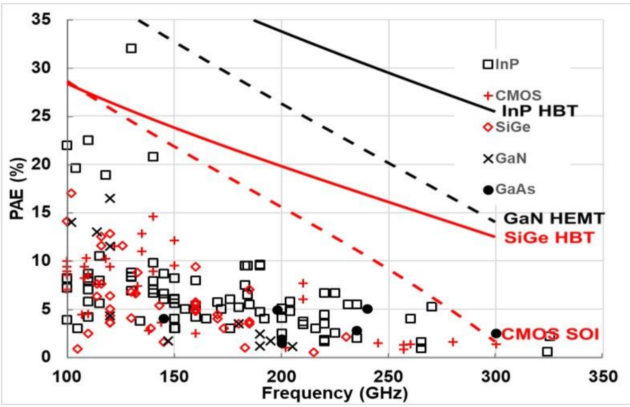


Fig. 5 - PAE versus frequency for PAs in the 100-300 GHz range. The solid and dashed lines indicate the theoretical bounds for the various technologies described in Table 2 as a funtion of frequency.

The PAE theoretical trend is illustrated in Fig. 5 as the solid and dashed lines and indicates that across this band InP HBTs hold significant promise for high PAE compared to other technologies. Above 100 GHz, recent work based on a class-B biased InP HBT has reported 32% efficiency and the design of this PA will be discussed in the next section [17]. Other recent work based on class-A InP HBT PAs has achieved higher output power (20 dBm) at slightly lower efficiency (20%) [16]. The InP HBT has consistently demonstrated the highest efficiency to 300 GHz due to the high fmax and, consequently, gain. Additionally, the InP HBT has a reasonable load-line matching condition for moderate power levels. This feature has been used to demonstrate wideband PAs above 100 GHz to cover waveguide bands [19][20].

For bands below 150 GHz, GaN, SiGe, CMOS have also been demonstrating promising results and could with future circuit and device development push beyond 20% PAE. Above 200 GHz, there remains no clear discrimination between the PAE of the various technologies at this point in time.

The significant gap between the theoretical bounds and the measured PAs raises substantial questions about the potential for practical high efficiency PAs and motivates the central theme of this paper. There are several explanations for the theoretical/measured gap. First, the gain near compression drops for most device technologies and, therefore, a maximum gain calculated from extrapolating the f_{max} is likely not accurate at high-frequency. For example, InP HBTs have different f_{max} based on the device load line. Second, modeling of transistors above 100 GHz is not extremely accurate due to the lack of direct model verification through load pull and other conventional PA design techniques. Effects, such as source/emitter inductance, impact the available gain. Additionally, passives are typically more lossy than anticipated due to the higher series resistance due to current crowding at high frequencies and the skin effect. Vias between metal layers or thru-substrate vias also play a dramatic role in the loss of passives above 100 GHz.

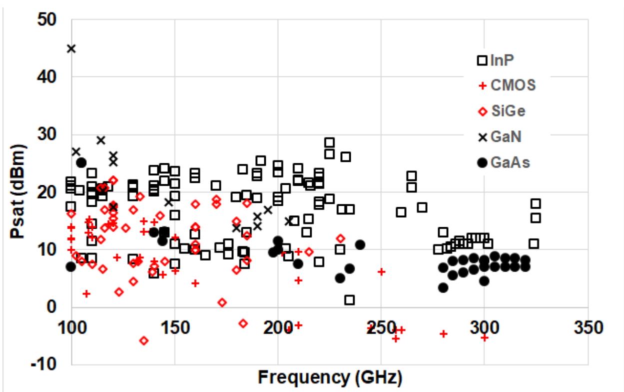


Fig. 6 - Psat versus frequency for PAs in the 100-300 GHz range

The saturated output power is plotted over the UmmW band in Fig. 6. The highest output power is demonstrated with GaN HEMTs up to 140 GHz and InP HBTs above 140 GHz. The rapid drop in power for GaN is due to the limited fmax of the technology. Optimizing GaN HEMTs for high-breakdown tends to also compromise the gate-drain capacitance parasitics that impact millimeter-wave performance. Recent work on N-polar GaN may offer new device physics for millimeter-wave operation [21]. Above 250 GHz, the InP HBT and GaAs mHEMT stand out as the only technologies that generate reasonable output power. SiGe has provided competitive performance in bands between 100 and 200 GHz [9]. Silicon technologies have to date only offered very limited power above 200 GHz. Recent work on Gmax-boosted approaches has pushed the output power towards 10 dBm with limited efficiency [22].

6. AN INP UMMW POWER AMPLIFIER

To demonstrate a high-efficiency PA design above 100 GHz, we have demonstrated circuit design techniques to maximize efficiency based on the 250-nm InP HBT process with f_{max} of approximately 600 GHz. As described previously, this InP HBT process offers a PAE as high as 45% due to 1) the high fmax and 2) load-line matching conditions close to 50 Ohms for an output power of 15 dBm.

Theoretically, the fmax is invariant to the choice of CE or CB configuration. While Common-Emitter (CE) amplifiers are conventionally used in PA design, we compare the MAG of CE and Common-Base (CB) amplifiers in the InP HBT process in Fig. 7 over a range of collector bias conditions at 140 GHz and 220 GHz. Generally, we observe that the gain remains relatively high in all cases but sharply reduces below a certain current density threshold. This substantial reduction in gain occurs when the K stability factor becomes imaginary. Over a range of collector biases between 0.1 mA/um and 3 mA/um (Imax), the CB provides 5dB higher than the CE transistor at both frequency bands. At 140 GHz, the CB provides higher gain close to the class-B biasing condition (Ic of zero) and the gain increases slightly as we shift to class A. The additional gain allows optimization for class-B operation in CB that would not be possible in CE.

Examining the 220 GHz operation, the CB gain drops 2 dB relative to 140 GHz. However, the gain of the CE transistor drops substantially (4 dB) as the desired collector bias shifts towards class-A. The gain at 220 GHz is 7 dB higher for the CB compared to the CE stage.

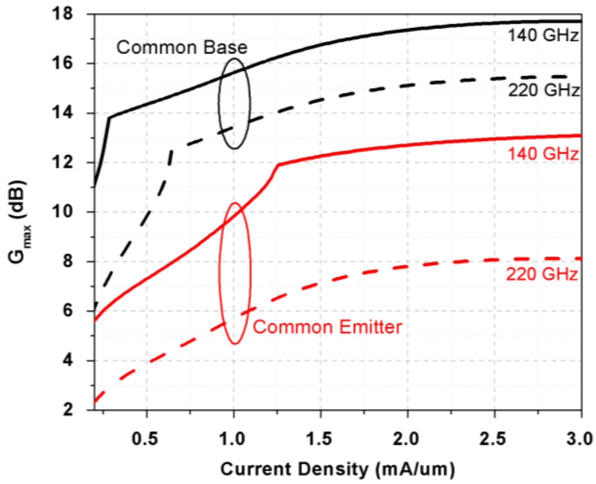


Fig. 7 - Common-base versus common-emitter for a constant collector-emitter voltage for a 0.25um InP HBT process.

The CB HBT provides higher MAG over the millimeter-wave band since the feedback parasitics in the CB amplifier are due to the collector-emitter capacitance, CCE, compared to the larger collector-base capacitance, CCB. Feedback current in the CB HBT is therefore much smaller than in the CE topology and the PA is unconditionally stable without additional stabilization. Base inductance typically impacts the stability of the CB configuration; however, the InP HBT process allows that the base can be directly connected to ground to eliminate any bypass capacitance requirement to produce an AC ground at the base node and the potential base inductance to connect to the bypass capacitor.

The InP HBT offers a physics-based scalable model that allows accurate load pull simulation of the CB device. A 4-finger by 4um (16 um total) CE or CB HBT emitter length produces a 100-Ohm load-line impedance for maximum gain and efficiency. Based on this transistor periphery, the output power, PAE, and gain at peak PAE for the CE and CB amplifier are plotted at 140 GHz in Fig. 8 as a function of the quiescent collector current (normalized by length) under the condition of fixed 2.5-V collector-emitter voltage. Note that the DC current differs significantly from the quiescent current as the PA shifts from class-B to class-A where the ratio of IDC/IQ approaches unity.

The peak output power is 16 dBm at class A for CE and drops to around 14 dBm under the class B bias. The CB configuration produces similar or slightly lower output power than the CE amplifier. Both configurations indicate identical PAE, which points to an invariance between CE and CB, that ranges from 46% in deep class AB and drops to 35% at class A. The gain of the PA at peak efficiency demonstrates that the CB offers significantly more gain than CE. Consequently, the CE has higher collector efficiency under the range of quiescent collector currents than CB but the higher gain of CB results in similar PAE.

The schematic of the pseudo-differential, common-base class-B PA is illustrated in Fig. 9 along with a chip microphotograph of the stage. The class-B voltage bias is provided to the emitter through the balun. The PA occupies an area of 0.4mm by 0.5mm to easily fit with the 140 GHz the grid spacing.

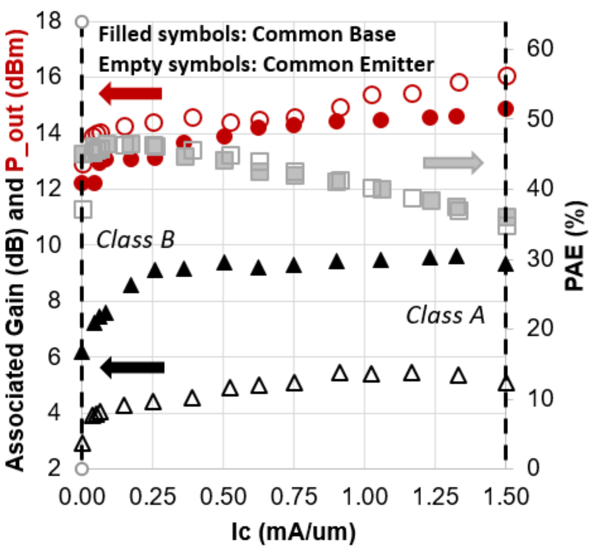


Fig. 8 - Gain, output power, and PAE at 140 GHz as a function of the quiescent collector current for the common-emitter and common-base amplifier for Vce = 2.5 V.

For CB HBTs, the ratio of impedance seen into the collector and at the output of the emitter is related directly to the power gain when IE ~ IC for sufficiently large. To increase the output power, two common-base stages are used as a pseudo-differential output PA. A low-loss sub-quarter wavelength balun was used at the input and the output to match the pseudo-differential PA at the input and output [7][23]. The balun places constraints on the maximum gain that can be realized from a CB amplifier. To provide 6 dB gain, we choose the impedance of emitter port to be close to 25 and for the 100-Ohm collector impedance. A series inductor at the output transforms the load impedance to 50 Ohms.

Fig. 10 plots the PAE and gain as a function of Pout at 130 GHz at a class-B collector bias current density of 203uA/um.

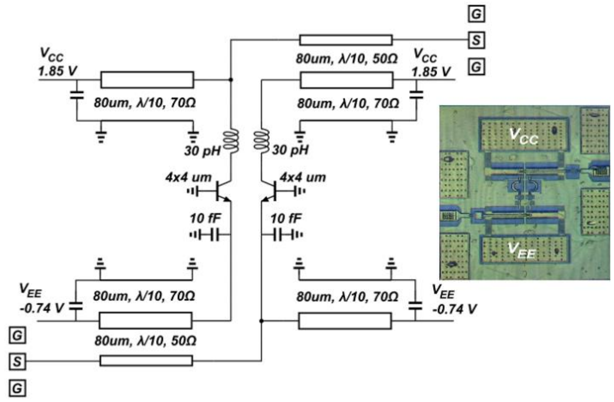


Fig. 9 – Pseudo-differential, common-base InP HBT class-B PA at 140 GHz. The area of the die on the right is 0.4 x 0.5 mm.

The peak PAE occurs at 0.3 dB higher output power. The PA exhibits a peak gain of 7 dB which occurs at Pout of 13 dBm with 1 dB of gain expansion due to the class-B biasing. The output loss of the matching network was measured through test structures to be around 1dB which raises questions about why the measured class-B gain is less than the maximum available gain indicated in Fig. 7. The explanation is partially explained when referring back to Fig. 8 which plotted the gain associated with the common-base device under the matching conditions for high efficiency. Here, the common-base device has only around 9 dB of gain near the class-B bias, corroborating the measured PA operating gain.

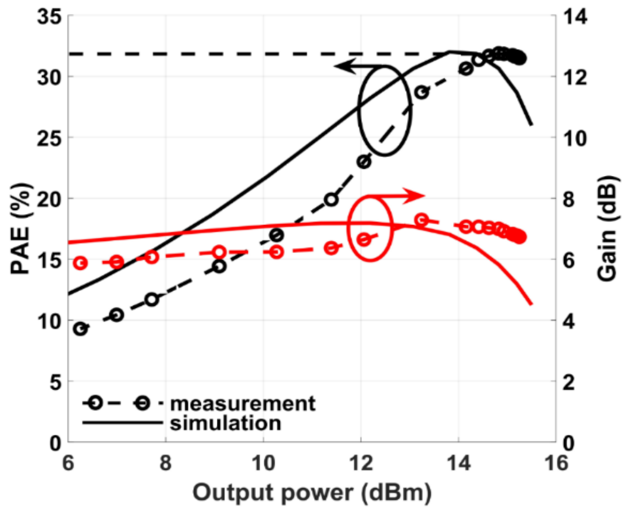


Fig. 10 - PAE and gain as a function of output power for VCC = 2.5 V at 130 GHz. From [15].

The PA achieves 32% peak PAE with 15.3-dBm saturated output power. The 1-dB power bandwidth covers 122 GHz to 146 GHz and is consistent with the measured 3-dB bandwidth from the S-parameters. The input power across the band is calibrated between 8 dBm and 8.7 dBm with variation due to the probe loss variation over the band.

To achieve higher output power, recent work has also investigated low-loss power combining from PA cells designed for high power. Fig. 11 illustrates the 8-way combined power amplifier, where each PA is based on common-base, class-A stages [24]. As opposed to the previous design where the power was combined through pseudo-differential stages (Fig. 10), power combining across two CB HBTs is performed to prevent a large emitter length device (lower left of Fig. 11). The output power combiner must be low loss for high PAE and compact for a small die area. While Wilkinson combiners are broadband, an 8:1 Wilkinson combiner requires 14 λ/4 transmission-lines with lossy, high-impedance lines. The proposed combiner is designed for a 50-Ω load without including the shunt inductive lines tuning Ccb. At the PA output, short 50-Ω transmission line sections (TL1) combine the outputs of the two 4×6um cells. At each consecutive level of combining, the characteristic impedance is divided in half, resulting in wider transmission lines and lengths that are minimized for smallest losses. A final impedance transformation from 12.5Ω to 50Ω requires a λ/4 line (TL3) having 35Ω characteristic impedance.

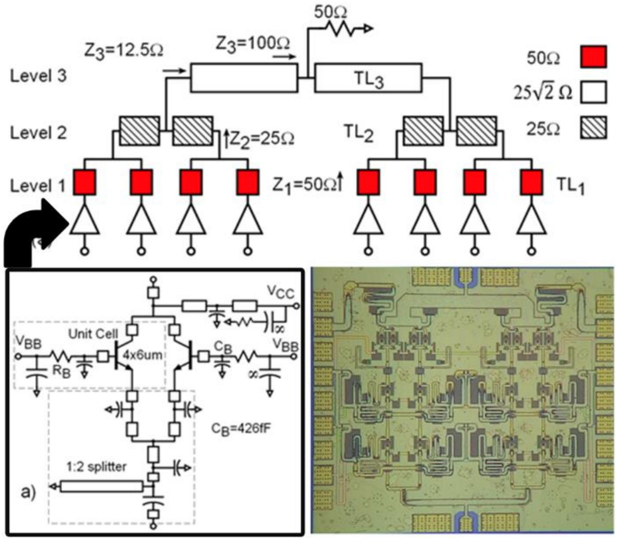


Fig. 11 - Power-combined, common-base InP HBT class-A PA at 140 GHz. The area is 1.23mm× 1.09mm.

Fig. 12 plots the PAE and gain as a function of Pout at 140 GHz at a class-A collector bias current density of 1.14mA/um. The three-stage PA has 23 dBm peak power with 17.8% Power Added Efficiency (PAE) and 16.5dB associated large-signal gain at 131GHz. At 131GHz, the small-signal gain is 21.9dB. The small-signal 3dB-bandwidth is 125.8-145.8GHz.

While the class-B PA and power-combined class-A PA offer state-of-the-art performance for high-efficiency and high-power, we continue to see opportunities to continue improving the PAE towards 40% at power levels exceeding 20 dBm.

While this work has not addressed trade-offs between power combining and device scaling in single-ended and differential PA designs, there remain significant research insights to be gathered about the losses and area efficiency of the various approaches. The power combiners in Fig. 11 occupy substantial area that might not satisfy the area constraints in a digital beam-former. Nonetheless, PA cells that are designed for a 50 Ohm load can be less risky than attempting to scale the device to meet similar power requirements.

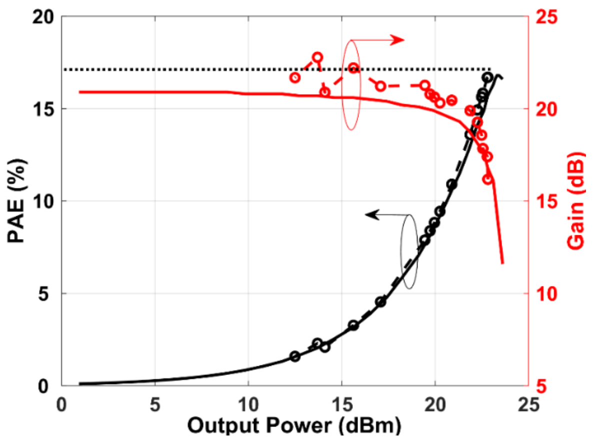


Fig. 12 - PAE and gain as a function of output power for VCC = 2.43 V at 140 GHz. From [28].

7. CONCLUSION

This paper has reviewed the requirements for power amplifiers in digital beam-forming arrays in frequency bands between 100 and 300 GHz. Efficiency will play a critical role in reducing the thermal load for front-end packaging due to high power density. We review the optimization of PAs in gain-limited operation and available device technologies above 100 GHz for PAs to construct PAE bounds on efficiency and compare recent published work to these bounds to demonstrate the potential for future research. Recent demonstrations of class-A and class-B power amplifiers in the 120-140 GHz range have set records for efficiency at 20% and 30%, respectively, which were substantial improvements over prior work. Further improvements in efficiency above 100 GHz are possible in all technologies and the frequency bands between 100-300 GHz may be as energy efficient as lower millimeter-wave bands while offering support of massive MIMO.

ACKNOWLEDGEMENT

This work was supported by Semiconductor Research Corporation (SRC) through the JUMP ComSenTer program. We also thank GlobalFoundries, Teledyne Scientific Corporation, and HRL through support of the DARPA MGM program for fabrication services.

REFERENCES

[1] M. J. W. Rodwell, "100-340GHz Spatially Multiplexed Communications: IC, Transceiver, and Link Design," Signal Processing Advances in Wireless Communications (SPAWC), Cannes, France, 2019, pp. 1-5.

[2] "CCIR Doc. Rep. 719-3,"Attenuation by Atmospheric Gases," ITU 1990."

[3] E. G. Larsson et al., ''Massive MIMO for next generation wireless systems,'' IEEE Communications Magazine, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[4] E. Torkildson et al., "Indoor Millimeter Wave MIMO: Feasibility and Performance," IEEE Trans. on Wireless Communications, vol. 10, no. 12, pp. 4150 – 4160, Nov. 2011.

[5] L. Kuai et al. "A N260 Band 64 Channel Millimeter Wave Full-Digital Multi-Beam Array for 5G Massive MIMO Applications," IEEE Access, August 2020.

[6] J. Buckwalter et al. "Prospects for High-Efficiency Silicon and lll-V Power Amplifiers and Transmitters in 100-300 GHz Bands," Custom Integrated Circuits Conference (CICC), Austin, TX, 2021.

[7] M. Bao, Z. S. He and H. Zirath, "A 100–145 GHz Area-efficient Power Amplifier in a 130 nm SiGe Technology," 12th European Microwave Integrated Circuits Conference (EuMIC), Nuremberg, 2017, pp. 277-280.

[8] H. Lin and G. M. Rebeiz, "A 110–134-GHz SiGe Amplifier with Peak Output Power of 100-120 mW," IEEE Transactions on Microwave Theory and Techniques, vol. 62, no. 12, pp. 2990-3000, Dec. 2014.

[9] S. Daneshgar and J. Buckwalter, "Compact Series Power Combining Using Subquarter-Wavelength Baluns in Silicon Germanium at 120 GHz," IEEE Transactions on Microwave Theory and Techniques, vol. 66, no. 11, pp. 4844-4859, Nov. 2018.

[10] D. Simic and P. Reynaert, "A 14.8 dBm 20.3 dB Power Amplifier for D-band Applications in 40 nm CMOS," IEEE Radio Frequency Integrated Circuits Symposium (RFIC).

[11] M. Rodwell et al. "InP Bipolar ICs: Scaling Roadmaps, Frequency Limits, Manufacturable Technologies," Proceedings of the IEEE, vol. 96, no. 2, pp. 271-287, Feb. 2008.

[12] M. Urteaga et al., "InP HBT Technologies for THz Integrated Circuits," Proceedings of the IEEE, vol. 105, no. 6, pp. 1051-1068, June 2017.

[13] J. Böck et al., "SiGe HBT and BiCMOS process integration optimization within the DOTSEVEN project," IEEE Bipolar/BiCMOS Circuits and Technology Meeting – BCTM, October 2015.

[14] R. A. Camillo-Castillo et al., "SiGe HBTs in 90nm BiCMOS technology demonstrating 300GHz/420GHz fT/fMAX through reduced Rb and Ccb parasitics," IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM) 2013.

[15] O. Inac et al., "45-nm CMOS SOI Technology Characterization for Millimeter-Wave Applications," IEEE Transactions on Microwave Theory and Techniques, vol. 62, no. 6, pp. 1301-1311, 2014.

[16] K. Shinohara et al., "220 GHz fT and 400 GHz fmax in 40-nm GaN DHHEMTs with re-grown Ohmic," in IEDM Tech. Dig., 2010, pp. 672-675.

[17] K. Ning et al., "A 130-GHz Power Amplifier in a 250-nm InP Process with 32% PAE," 2020 IEEE Radio Frequency Integrated Circuits Symposium (RFIC) June 2020.

[18] A. Ahmed et al., "A 140GHz power amplifier with 20.5dBm output power and 20.8% PAE in 250-nm InP HBT technology," IEEE/MTT-S International Microwave Symposium (IMS) June 2020.

[19] Z. Griffith et al. "A 115–185 GHz 75–115 mW High-Gain PA MMIC in 250-nm InP HBT," 2019 49th European Microwave Conference (EuMC).

[20] Z. Griffith, M. Urteaga and P. Rowell, "A 140-GHz 0.25-W PA and a 55-135 GHz 115-135 mW PA, High-Gain, Broadband Power Amplifier MMICs in 250-nm InP HBT," 2019 IEEE MTT-S International Microwave Symposium (IMS), Boston, MA, USA, 2019, pp. 1245-1248.

[21] B. Romanczyk et al., "N-Polar GaN-on-Sapphire Deep Recess HEMTs With High W-Band Power Density, IEEE Electron Device Letters, vol. 41, no. 11, November 2020.

[22] H. Bameri and O. Momeni, "An Embedded 200 GHz Power Amplifier with 9.4 dBm Saturated Power and 19.5 dB Gain in 65 nm CMOS," IEEE Radio Frequency Integrated Circuits Symposium (RFIC), June 2020.

[23] H-C Park et al., "Millimeter-Wave Series Power Combining Using Sub-Quarter-Wavelength Baluns," IEEE Journal of Solid-State Circuits, vol. 49, no. 10, pp. 2089-2102, October 2014.

[24] A. Ahmed et al., "A 200mW D-band Power Amplifier with 17.8% PAE in 250-nm InP HBT Technology," 2020 15th European Microwave Integrated Circuits Conference (EuMIC).

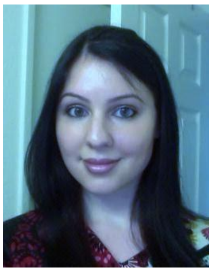
AUTHORS

James F. Buckwalter received a B.S.E.E. degree with honors in electrical engineering from the California Institute of Technology (Caltech), Pasadena, in 1999, M.S. degree from the University of California, Santa Barbara in 2001, and Ph.D. degree in electrical engineering from Caltech in 2006. He is currently a professor of electrical and computer engineering with the University of California - Santa Barbara (UCSB), Santa Barbara.

Mark J. W. Rodwell holds the Doluca Family Endowed Chair in Electrical and Computer Engineering at UCSB and directs the SRC/DARPA ComSenTer Wireless Research Center. His research group develops high-frequency transistors, ICs and communications systems. He and his collaborators received the 2010 IEEE Sarnoff Award, the 2012 Marconi Prize Paper Award, the 1997 IEEE Microwave Prize, the 2009 IEEE IPRM Conference Award, and the 1998 European Microwave Conference Microwave Prize.

Kang Ning (S'14-M'20) received a B.S. degree from the School of Physics Science and Technology, Lanzhou University, Lanzhou, China, in 2012, an M.S. degree in electrical and computer engineering from the University of Minnesota at Twin Cities, Minneapolis, MN, USA in 2015 and Ph.D degree in electrical and computer engineering from the UCSB in 2020. He is currently working at Mixcomm Inc.

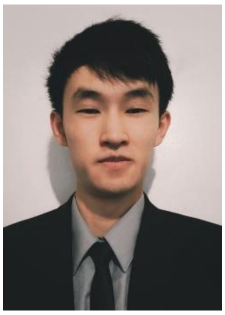
Ahmed Ahmed received B.S., and M.S. degrees in electrical engineering from Cairo University, Cairo, Egypt, in 2012, and 2015 respectively. He received a Ph.D. degree in electrical and computer engineering at the UCSB in December 2020. He is currently working for Marki Microwave Inc.



Andrea Arias-Purdue received B.S. and M.S. degrees in physics from California State University, Fresno. She has 10 years of experience in high-power device and IC design. She is a Member of Technical Staff at Teledyne Scientific, where she serves as PI or PM for various government-funded efforts. Ms. Arias-Purdue is also currently pursuing her PhD at UCSB focusing on device and amplifier MMIC design automation for high power and efficiency above 100 GHz.

Jeff Chien received B.S. and M.S. degrees in EECS and EE from National Tsing Hua University in 2014 and 2015, respectively. During 2015 and 2018, he was with Airoha Technology Corporation. He is now pursuing his Ph.D. at UCSB on the design of sub-THz transmitters and power amplifiers.

Everett O'Malley received a B.S. degree in electrical engineering and physics from Northeastern University in 2018. He is currently a Ph.D. researcher at UCSB. His current work focuses on the design of high-efficiency power amplifiers using embedding networks.



Eythan Lam received a B.S degree in electrical engineering from UCSB, CA, USA in 2020. He is currently working towards a Ph.D. degree at UCSB. His current research focus is on high-efficiency power amplifier design and digital pre-distortion.

SELF-CONFIGURING ASYNCHRONOUS SLEEPING IN HETEROGENEOUS NETWORKS

Ali Medlej[1], Eugen Dedu[1], Kamal Beydoun[2], Dominique Dhoutaut[1]
[1]FEMTO-ST Institute/CNRS, Univ. Bourgogne Franche-Comté, Montbéliard, France, [2]L'ARICoD Laboratory, Faculty of Sciences I, Lebanese University Beirut, Lebanon

NOTE: Corresponding author: Ali Medlej, ali_ghasswan.medlej@univ-fcomte.fr

Abstract – *Nowadays, the heterogeneous wireless nano-network topology becomes a need for diverse applications based on heterogeneous networks composed of regions of different node densities. In Wireless Nano-networks (WNNs), nodes are of nano-metric size and can be potentially dense in terms of neighbouring nodes. Nano-nodes have limited resources in terms of processing, energy and memory capabilities. In nano-network(s), even in a communication range limited to tens of centimeters, thousands of neighbours can be found. We proposed a fine-grained duty-cycling method (sleeping mechanism), appropriate to nanonodes, which aims to reduce the number of receptions seen by a node during data packet routing. The present study reveals the usefulness of implementing the sleeping mechanism in heterogeneous networks, as well as configuring a dynamic awaken duration for nodes based on a density estimation algorithm. We also proposed an algorithm that helps in increasing the reliability of the packet received by the destination node.*

Keywords – Density estimation, heterogeneous nano-network, sleeping mechanism, routing

1. INTRODUCTION

Nano-network(s) are made up of tiny nodes, called nanonodes, of a nanometric size. These nodes use electromagnetic waves from the THz band (0.1–10 THz) [1] for their communication. Nano-nodes possess sensors, actuators, a processor, and a memory. They can move and communicate with each other [2]. These nodes are limited in their computing capabilities, storage, and energy. Therefore, they need to collaborate with other nodes to fulfil their tasks [3].

1.1 Modulation and routing techniques

Given the very limited energy of nanonodes, a modulation technique based on femtosecond pulses for terahertz communications called Time Spread On-Off Keying (TS- OOK) was proposed by Jornet and Akyildiz [4]. TS-OOK is based on femtosecond-long pulses where communica- tion data is sent using a sequence of pulses interleaved by a randomly selected constant duration (Fig. 1). TS-OOK uses an electromagnetic pulse of duration T_p to transmit a bit "1", and silence (no transmission) for bit "0". The time between two consecutive bits is T_s. Due to hardware and power constraints, the spreading ratio $\beta = T_s/T_p$ can be very large. Using T_p = 100 femtosecond, the total available bandwidth in the network is very high, in the order of terabit per second.

In very dense nano-network(s), a nanonode cannot frequently update its neighbourhood list. Also, the communication range is short and the desired destination node is often beyond its communication range. Therefore, it is necessary to have a routing protocol that performs the tasks of data routing between nodes. A traditional addressing scheme like IP is useless. An alternative addressing way must be used. For that, a spatial addressing routing protocol that forwards messages *in the direction* of a destination could be used.

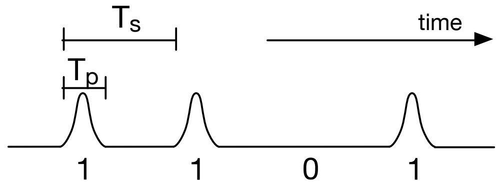


Fig. 1 – TS-OOK modulation.

Stateless Linear-path Routing (SLR) [6] is a spatial addressing and routing protocol we use in our evaluation. It implements a coordinate-based routing, in which data packets are routed in a linear routing path. The SLR protocol has two phases: (1) initial/addressing phase and (2) routing phase.

1.2 Applications and node limitations

Nano-network(s) can be potentially used in a wide range of applications and fields such as: (1) health field (e.g. monitor human vital signs, drug delivery systems), (2) military field (e.g. border control to prevent any penetration), (3) agricultural field (e.g. monitoring the temperature, humidity, and water level of plants), (4) telecommunication sector [5]. In such applications the network topology is often heterogeneous. A heterogeneous nano-network is a network composed of zones of different densities. Nowadays, many applications use this type of networks. For example, in agricultural applications, some areas have high nodes' density while others have a small density due to the varied geographical terrains. Also, the human body is considered as a natural heterogeneous network because of its internal and external structure.

An even more severe limitation (and a main difference from traditional wireless networks) lies in the inability of a single node to process the whole channel capacity. In the typical networks, a node may always listen to the channel and take whatever packet concerns it. There is no way for a tiny, energy constrained nanonode to sustain Tb/s traffic. This translates into a much smaller effective channel capacity, as all nodes in an area may saturate their reception capability way before saturating the channel itself. A typical solution to this problem is to let the nodes sleep from time to time. A traditional sleeping mechanism cannot be applied in a nano-scale network. Our proposition differs significantly from the previous schemes that make nodes sleep or awaken for periods longer than a packet. To cope with the very specific characteristics of nanonet- works, our idea is to make the node inactive for a fraction of T_s (the time between two consecutive bits). We then make it awake in order to receive the intended bit. Such a mechanism aims to preserve node resources (CPU, mem- ory, energy), therefore extending network lifetime.

In this paper: (1) we show that the sleeping mechanism works in a heterogeneous nano-scale network, (2) we propose a method to automatically tune the awaken duration of a node depending on the average density of neighbours (awakenNodes), (3) we propose an algorithm for packet retransmission at the destination zone, and (4) we test this method and show its impact on increasing received packets by the destination node.

The rest of this paper is organized as follows. Section 2 introduces applications used within heterogeneous networks. Section 3 describes the algorithms used in the paper. Packet retransmission algorithm is explained in the Section 4. Evaluation via simulations takes place in Section 5. Finally, we conclude in Section 6.

2. APPLICATION FIELDS

Different approaches in the literature focus on designing and developing routing protocols for heterogeneous electromagnetic nanonetworks. Nowadays, the nano-scale communication technology enters into many important sectors such as health, military, and agriculture.

2.1 Health sector

Heterogeneous networks pervade many applications where nodes are distributed depending on the environment conditions. Due to their nanoscale size, nanonodes can be deployed at different scales [7] inside the human body and across a variety of environmental contexts. Nanonode uses are not limited to sensing and monitoring human vital signs, but it may cover some other operations when needed.

An example of a medical application that can be assigned to nanonodes and needs high reliability and accuracy is *drug delivery* (Fig. 2). Heterogeneity resides in the human

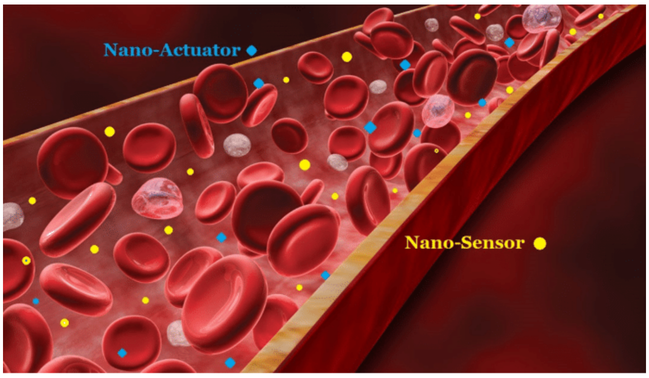


Fig. 2 – Drug delivery in human body.

body through the variation in the size and shape of the human organs [8]. Nanomedicine may be defined as the use of nano-devices and nanostructures for monitoring, repairing or controlling human biological systems at the molecular level [9]. The nanonodes spreading in blood vessels can monitor the glucose level, and at the same time release the insulin to regulate the glucose level. The risk lies when the patient forgets to take his medicine especially those related to the control of certain levels of enzymes in the blood (as in the example above). With this type of technology, the patient becomes more able to con- trol his health troubles.

This heterogeneity of the human body poses a major challenge to the routing protocol by traversing areas of different densities.

2.2 Agriculture sector

Agriculture is an important source of livelihood in most parts of the world. Wireless nano-sensor(s) have been used in modern agriculture and farming (Precision Agriculture (PA)), defined as the techniques of applying farming parameters and resources for production optimization, and reducing human effort [10].

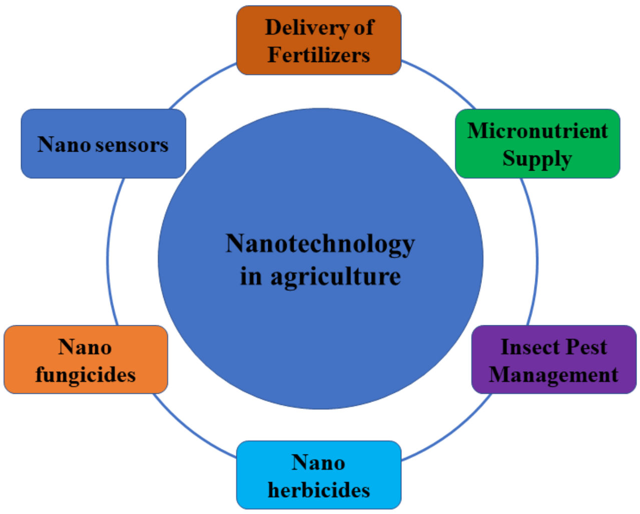


Fig. 3 – Nanonetwork technology impact in agriculture.

To improve productivity (e.g. detection of plant viruses, soil humidity degree), a farmer needs to monitor numerous parameters (Fig. 3). Wireless nano-sensors need to be used to realize the vision of intelligent agriculture. One can deploy autonomous nano-sensors around the plants to monitor soil condition and plant viruses, which gives them all the information about the environmental condi- tions of the plants using a simple portable device.

WNSN technology will contribute in generating the tools for establishing a real-time plant monitoring system, composed of a chemical nano-sensor merged with plants. Chemical nano-sensor nodes are miniaturized machines that interact with the environment to collect chemical compounds disseminated by plants. Micro-gateways can interconnect data collected from the nanonetwork to the external network, and to the decision officer of the analytical laboratory [11].

2.3 Civil engineering sector

Developments in nanotechnology can benefit construction engineering by enabling the practical deployment of structural condition monitoring systems for large civil engineering systems (Fig. 4). Nanonodes can be integrated into a composite material that can provide information on its performance and environmental conditions by monitoring structural loads, temperature, and humidity. It can be used for the construction of smart buildings [12]. Also, they could be coated into bridges, as prevention monitoring against degradation and cracking in order to alert the decision-making authorities well before the damage is de- tectable by human inspectors [13].

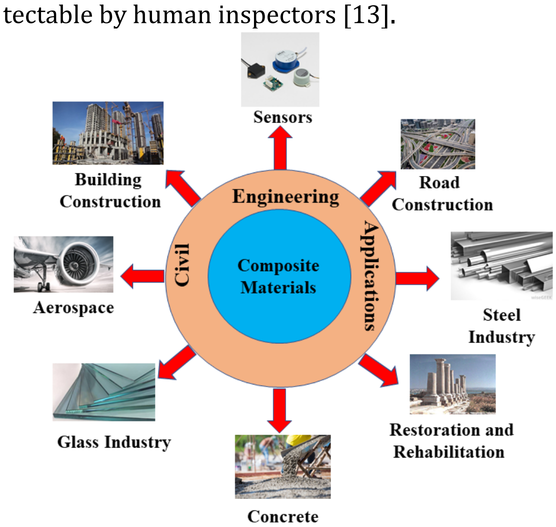


Fig. 4 – Smart composite materials based on nanotechnology, for civil engineering applications.

Nanonodes could also be embedded into roads and structures to allow engineers to monitor deterioration and cracking, thus saving physical intervention. Road sensor networks could gather and provide data to transport operators to better control and detect congestion and incidents.

2.4 Environment sector

Environmental safety is one of the sectors where nanotechnology may have a major impact. Through the installation of nano-sensors in high density public locations (e.g. hospitals, airports, and restaurants), authorities can trace the circulation of viruses and notice how various types of people are affected. Nanosensor networks could also be utilised to monitor the environment, such as pollution and gas emission.
Water is the lifeblood of the world, hence the importance of monitoring its quality and safety. We can benefit from nano-sensors in detecting bacteria, diseases and other harmful infectious agents, especially when improvement is made close to the point of use at the household. Nanosensors can also be used in flood-prone rivers near residential places. These sensors can measure the water level in rivers, and send the information to the control and decision room in real time (Fig. 5).

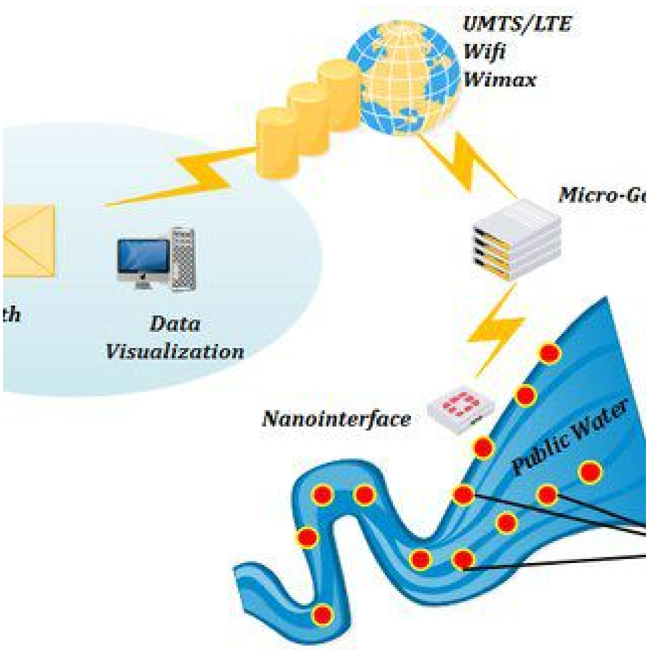


Fig. 5 – A water monitoring architecture.

To summarise, numerous applications use heterogeneous wireless networks. The biggest challenge remaining is the design of the routing protocols, which will face the challenges of different node densities in several regions, and in this context we focus on the influence of the sleep mechanism on the delivery of data packets to their desired destination. This is the focus of the evaluation section in this article.

3. BACKGROUND

In this work, we target potentially very dense, multi-hop nano-network(s). The contributions of this paper are presented later, but they are built upon a collection of protocols already independently designed for those dense nano-networks. In this section we give the necessary background concerning these protocols that we are using together: the routing protocol (SLR), the nano-sleep mechanism itself, and the density estimation protocol (DEDeN).

It is important to note that the choice of the routing protocol here is mostly inconsequential to the contribution. In this paper we do not present a new routing protocol. In regard to the very specific concurrent transmission behaviour of wireless nano-network(s), we instead present a mechanism to reduce the load on each node. This mechanism can be adapted to various routing protocols and here we choose SLR for the sake of convenience.

3.1 SLR protocol

Stateless Linear-path Routing (SLR) is a spatial routing/addressing protocol. It implements a coordinate-based routing, in which data packets are routed in a linear routing path. Coordinates are defined as an integer number of hops from special nodes called anchors (Fig. 6). All nanonodes are assumed to be placed in a cubic space, where anchors' nodes are placed at the vertexes (two anchors for a 2D area / three anchors for 3D area).

The SLR protocol has two phases: (1) addressing/initial phase and (2) routing phase.

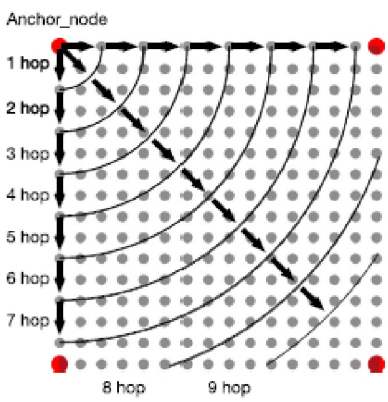


Fig. 6 – SLR addressing phase.

In the initial phase, addressing, as shown in Fig. 7, two anchor nodes (placed at the vertexes of 2D network) broadcast a packet (beacon) to the whole network. Beacon mes- sages include a hop distance field initialized to 0, which increments with each retransmission. The node coordinates are the hop counter in those beacons and correspond to the distance to the anchors. This phase will be performed only once at the network deployment.

In the routing phase (Fig. 7), packets contain the SLR address of the sender and receiver. A greedy approach is used to route the packets to their intended destination. When a node receives a message, it checks if it is located on the path from the source to the destination. If and only if it is the case, the node forwards the message. This check uses basic integer computation appropriate for the low nanonode computation capabilities.

Note that because of the way the coordinates are defined,they can be shared by multiple nodes (which belongs to the same SLR zone). Consequently, SLR is a protocol that routes a packet to the correct zone instead of a specific node.

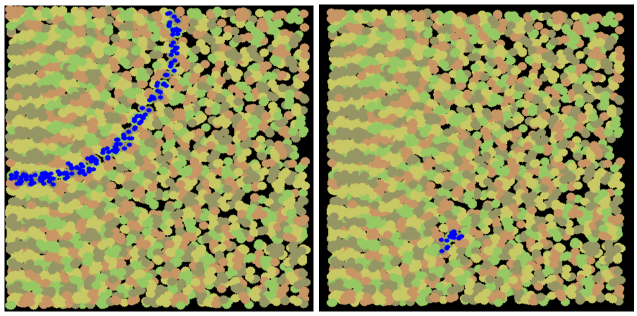


Fig. 7 – SLR initial phase (left) and SLR routing phase (right).

3.2 Sleeping mechanism

We start be emphasizing that the **nano-sleep** we present in this section is different in both means and objectives from the sleep mechanism (often call **duty-cycle**) that we commonly find in traditional networks.

In traditional networks, making a node sleep and wake up in defined intervals aims at preserving energy and thus increasing the life span of the network. A node always wakes up for a duration long enough to allow the reception of one or more frames. It processes the received packets and then goes back to sleep by turning off most of its reception and processing capabilities. By doing so, it preserves a substantial amount of energy. As the packets are sent sequentially over the channel, the longer the awaken duration, the larger the number of packets received. We could also say that while it is awake, a node has to be able to process the full throughput of the network. It then completely ignores packets sent while it is sleeping.

Our proposed nano-sleeping mechanism differs from those used in a macro-scale network by its fine granularity, asynchronicity, and decentralization. The main problem it solves is the potential overwhelming of a node when too many packets coexist on the channel. With TS-OOK, multiple nodes targeting different receivers can indeed transmit at the same time, and there is no hard limit to how many packets can **concurrently** coexist on the channel. Because of that, using a traditional awaken duration may cause a node to be overwhelmed when the number of packets concurrently transmitted exceed its decoding capabilities. Nano-sleeping is designed to tackle this problem, by allowing a node to concentrate on a small number of packets while completely (and safely) discarding the others.

A nanonode does not stay awake for the duration of one or several packets, but for a much shorter duration, a fraction of the T_s value. In our proposed sleeping mechanism, all nodes have the same awake-sleep cycle, equal to T_s. Inside the cycle, all the nodes have the same

awake duration (or percentage of T_s) , but the beginning of the awake interval is different for each node, and is randomly deter- mined. The mechanism ensures that if a node is able to pick the first bit of a packet, it will be able to pick all the following ones. More details on this sleeping mechanism are described in [14].
Fig. 8 depicts an example, where receiver nodes Recv1, Recv2 and Recv3 wake up at different times, but for the same duration. Recv1 and Recv2 are able to pick the bits from flows 1 and 2, as they arrive when they are awake. Recv3 is able to pick bits only from flow 3.

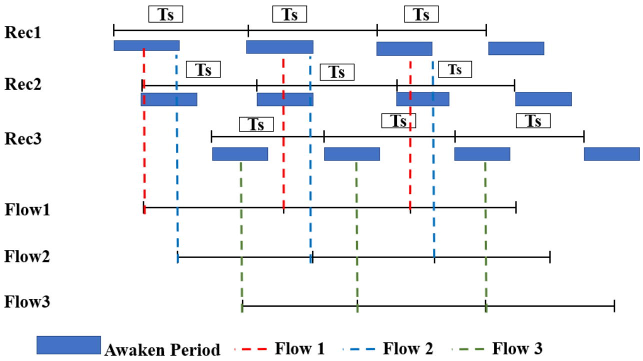


Fig. 8 – Sleep mechanism with three nodes and three flows.

The very high frequency of awake and sleep transitions make it significantly different from a traditional duty-cycle. Fully turning on and off processing capabilities at such a speed may not be technically achievable, but that is not the main point of our protocol anyway. Instead, we just consider that nodes do not process bits (pulses in TS-OOK) outside of their awaken duration.
As nodes do not turn off most of the hardware, the energy saving is limited (but still definitely exists, given that ignored frames are not processed and potentially retransmitted).
As we will see later, the benefit of this mechanism lies in an increase of the overall **effective** channel capacity. This benefit becomes really significant as the local node density increases and nodes can have individual lower and lower awaken durations.

3.3 Density estimator algorithm

Because nodes do not have much memory or processing power, elaborate routing protocols that try to find optimum forwarder(s) cannot work well if at all. Hopefully, we can improve routing protocols by using the local density of the network to limit their retransmission rate. We need a way for the nodes to discover by themselves how many neighbours they have.
A traditional solution to this problem is that each node sends "HELLO" packets. All nodes would only need to maintain a list of the neighbours from which they have received an HELLO packet to know the local density. This solution is simple and efficient if the density is low, but it becomes prohibitively costly in resources consumption as the density increases.

Density Estimator for Dense Networks (DEDeN) is an algorithm tailored to wireless nano-networks. It allows a node to estimate the number of its neighbours (also called node degree, node density, neighbour density, neighbourhood density, local density, or simply density).

DEDeN works by making nodes transmit an initialization message, and all nodes that receive it start the discovery process. With a very low (but known in advance) probability, they may start sending small beacons. As the probability is known, nodes receiving those beacons can easily compute an estimated local density. If the confidence in the estimation in not high enough, nodes start another round, with a higher beaconing probability. The confidence increases with the number of packets received and the probability to transmit.

The confidence and the error range of the estimation can be adjusted to the requirements of the user with a predictable overhead. Depending on how it is initiated, DEDeN enables a unique node or all nodes to estimate the neighbourhood density. The algorithm may be executed each time this neighbourhood density is needed.

For additional information about the density estimator algorithm, refer to [15].

4. RETRANSMISSION ALGORITHM AT THE DESTINATION ZONE

As explained in the previous section, the proposed sleeping mechanism should reduce the congestion and preserve node resources consumption. When a packet arrives at the destination zone, if the destination node is asleep, it will lose the data.

It is important to note that the proposed sleeping mechanism was implemented and tested in a homogeneous network. In this paper we are now evaluating this protocol in an heterogeneous network (areas of different nodes density). All nodes have the same awake-sleep time T_s, but they are different in their awaken duration and awaken starting time.

We propose an algorithm that increases the destination node chance of receiving a packet even if it was asleep when the packet reached the destination zone. The algorithm is to be used only by nodes at the destination zone. No matter the network density, it never saturates the radio channel and does not require much memory or computations. The only memory needed is the buffer to store the received packet to retransmit it at the end of the node awaken duration. A summary of the packet retransmission algorithm is presented in Algorithm 1.

Algorithm 1 Packet retransmission algorithm executed by all nodes at the destination zone except the destination node.

```
alreadyseen = false
waitingTime
wT1ts = aD - (pcktrecp - aS)mod T_s
wT2ts = - (aD - (pcktrecp - aS))mod T_s
if packet type is data and reach the destination zone then
    if Sleeping mechanism is used then
        if packet !alreadyseen and the received node is not the destination node then
            if pcktrecp mod T_s ≥ aS then
                waitingTime = wT1ts
            else pcktrecp mod T_s < aS + aD - T_s
                waitingTime = wT2ts
            end if
        end if
    end if
end if
```

We recall that this algorithm works when the packet reaches the destination zone. The algorithm works on homogeneous and heterogeneous networks, with equal or different awaken durations. In this algorithm we took into consideration the case when the awaken duration spans over two time cycles. We describe the algorithm as follows:

- *waitingTime*: Represents the time the node should wait before packet retransmission at the end of its awaken duration.
- *wT1ts*: Formula to calculate the node waiting time if the awaken duration range is 1 T_s.
- *wT2ts*: Formula to calculate the node waiting time when its awaken duration spans for 2 T_s.
- *aD*: Represents the node awaken duration.
- *aS*: Represents the node awaken starting time.
- *pcktrecp*: Represents the time when the node receives the packet.

Fig. 9 depicts an example of a packet retransmission at the destination zone. In our proposed algorithm, all the nodes have the same awake-sleep cycle T_s. Inside the cy- cle, all the nodes are different in their awake duration aD (or percentage of T_s) , also the beginning of the awake in- terval for each node aS are randomly determined. When a node receives the packet, the algorithm checks whether the node is not the destination node. If this is the case, the algorithm will be applied.

Node 3 represents the destination node. Node 1 receives the packet, and since it is not the destination node, it executed the algorithm. The algorithm computes the waiting time to retransmit the packet at the end of the node awaken duration. By coincidence, nodes 2 and 4 are awake during the packet retransmission by Node 1. It happens that a node can receive several copies of the packet, which is the case for Node 4. This is due to the simultaneous awake of the node with the retransmission. Finally, the figure shows that the destination node awaken time does not overlap with any of the other nodes. Node 4 retransmits the packet at the end of his awaken duration, and it is received by the destination Node 3. Therefore, in the absence of the retransmission algorithm, the destination node would not have received the packet in the case where it was in sleep mode.

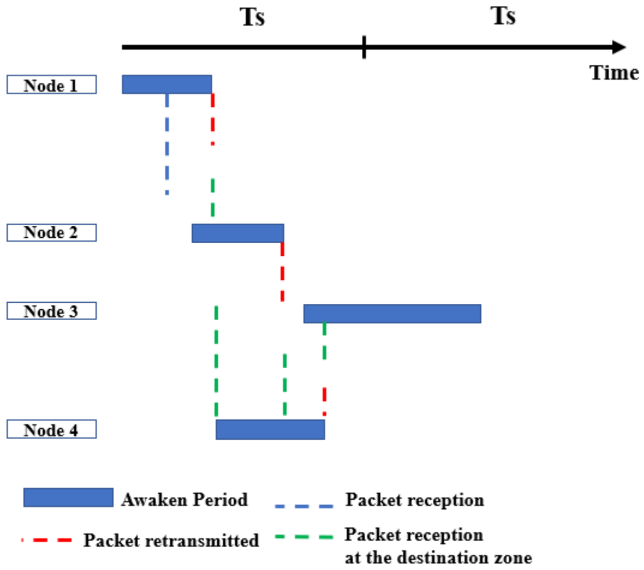


Fig. 9 – The effect of the retransmission algorithm at the destination zone.

5. EVALUATION

In this section, we evaluate our proposed ideas through simulations since a detailed analytical study is not possible. BitSimulator [16] is used in our simulations due to its characteristics.

BitSimulator has been designed to allow simulations of applications and routing protocols in wireless nano-networks. It differs from other network simulators by its accuracy as it can simulate every bit of payload and their potential collisions. It also targets ultra-dense nano-networks (hundreds of thousands of nodes). It is highly optimized for speed and accompanied by a VisualTracer, an interactive visualization and analysis tool that greatly helps in dense and complex scenarios. Technical details and information about full reproducibility of our results are provided on a separate website[1].

In our previous publication [14], we proposed a novel sleeping mechanism in a homogeneous nano-network which is totally different to the traditional sleeping mechanism applied in ad hoc networks. We define a homogeneous network as a network in which node density is constant. Therefore, all nodes will have the same awaken duration. In this kind of network, the nano-sleep mechanism

[1] `\http://eugen.dedu.free.fr/bitsimulator`

has been proven to work, and has contributed to improving the nodes' energy consumption as well as increasing network lifetime.
Nano-networks have different applications in various fields, extending from environment monitoring, industrial manufacturing, and agriculture to an enormous number of applications in medicine (like drug delivery, diagnostics and surgical operations).
Nowadays, the health sector is one of the most important sectors in which nanotechnology is involved. A nanonetwork can collect vital patient information and provide it to computing systems, providing more accurate and efficient health monitoring. These real applications will mostly be implemented in a heterogeneous nanonetwork due to its structural nature. This is why in this paper it was necessary to understand and validate the behaviour of the routing protocol used in places where nodes have different densities.

5.1 Network scenario

In our simulations, the network topology consists of a heterogeneous network as a 2D area of size 6 mm x 6 mm, shown in Fig. 10. It is composed of three rectangles of size 6 mm x 2 mm, with node average densities of: (rectangle 1: 160, rectangle 2: 80, and rectangle 3: 60). Nodes are always placed randomly except anchors, sources, and destinations nodes. Source nodes are placed at the bot- tom of the network near the edges.

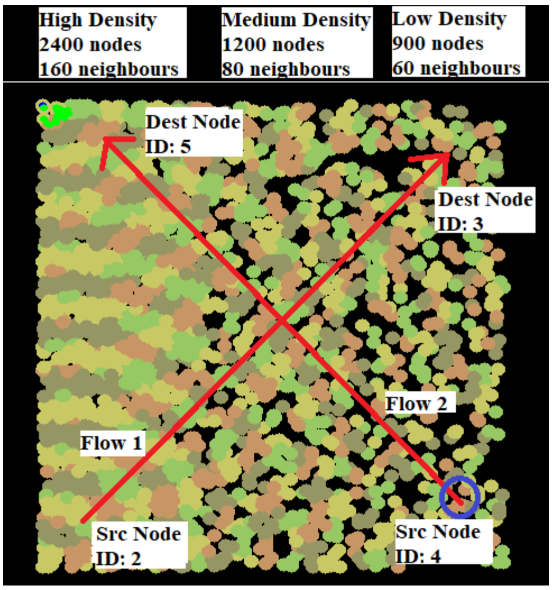


Fig. 10 – The network topology used, with 2 flows and 3 rectangles of different densities.

Given the environment size and the communication range (0.5 mm), the furthest possible receiver is in the corner at $\sqrt{2} * rectangle\ side/2 = 4.2$ mm away from the source. It means that a packet from the source needs at least 17 hops ($(4.2/0.5) * 2 = 17$) to reach the furthest possible node. Table 1 lists the parameters used in the simulations.

Table 1 – Simulation parameters

Size of simulated network	6 mm * 6 mm
Size of rectangles	6 mm * 2 mm
Number of nodes	4500
Communication radius	500 µm
Hops to reach the furthest node	17
β (time spreading ratio)	1000
T_p	100 fs

There are 2 flows. The source nodes are at the bottom right and bottom left of the network, while the destination nodes are at the top left and top right. Flow 1 traverses the network from left to right through different nodes' density areas (high, then medium, then low), while is the opposite for flow 2. According to the previous network settings, packet routing will be somewhat complicated and challenging.
The sleeping mechanism is applied to all nodes. To avoid random effects, each point in the following figures represents the average of 10 simulations, each with a different random generator seed to set the beginning of the sleeping period for nodes. In each simulation round, the awaken duration percentage and the number of neighbours are changed, while the other parameters are kept identical.

5.2 Comparison metrics

We analyse the sleeping mechanism efficiency using three metrics:

- *Packet reachability*: The number of packets that have reached the destination node. If several copies of the same packet reaches the destination, only one packet is counted.
- *Total number of sent packets*: The total number of sent packets in the network, for instance in a multi-hop transmission a same packet is sent by several nodes (routers), hence it is counted several times.
- *Awaken duration*: The interval of time where the node is awake. After this duration, the node goes back to sleep for the rest of the duty cycle (T_s). In BitSimulator, through the command line we can specify the percentage of awaken duration as a variable, or the average density of neighbour nodes (*awakenNodes*) via DEDeN that calculates it automatically.

5.3 Same awaken duration for all nodes

We recall that the time between two consecutive bits is $T_s = T_p * \beta$.. To determine an awaken percentage for every node in the network (e.g. 20% is equivalent to 20 000 fs) means that all nodes will awake for this percentage in a time duration equal to T_s. Inside the cycle, all the nodes have the same awaken duration (or a percentage of T_s), but the beginning of the awake interval is different for each node and is randomly determined.

Fig. 11 shows the average of packet reachability, i.e. at least one packet succeeding in reaching the destination node. We notice that packets are unable to reach the destination for an awaken duration ranging from 5 to 20%. This is due to the insufficient number of awaken nodes especially in low density areas, the flow propagation stops at the beginning. Recall that we are testing in a heterogeneous network, where areas' densities are different.

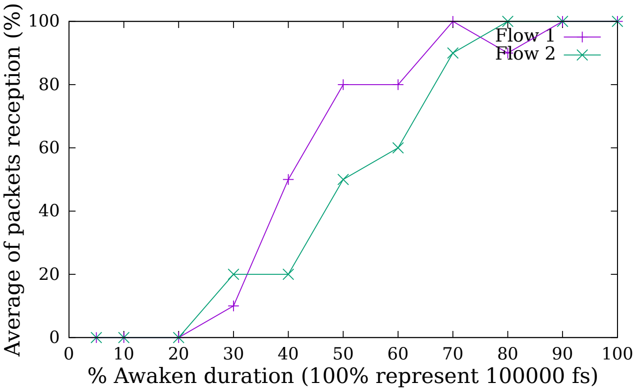


Fig. 11 – Percentage of arrived packets depending on awaken duration, for 2 flows.

Increasing the awaken duration immediately increases the average packet reachability. An awaken duration of 70 000 fs allows all packets from flow 1 to be received. Recalling that due to the sleeping phenomenon, it is possible for the packet to arrive while the destination node is asleep, and this results in the packet being lost. This is what the curve shows at 80 000 fs (flow 1). For flow 2, an awaken duration of 80 000 fs allows a total packet reception. *Note that 100% of awaken duration means that all nodes in the network are awake all the time.*

Given that the destination is far away from the source, the packet transmission between nodes occurs by hops, so a packet is sent several times, by several nodes in the path. The number of packets sent varies with the node awaken duration. Fig. 12 shows that the number of packets sent is very close to 0 at the beginning for both flows, due to the low number of awake nodes. Increasing the awaken duration percentage raises the number of awaken nodes, which results in an increasing number of packets sent too. This number reaches a value of 238 (flow 1) and 179 (flow 2) where the awaken duration is 90%.

5.4 Different awaken durations for nodes, based on local density

The sleeping mechanism dispatches the load of sent packets among neighbouring nodes. Therefore, taking into account the neighbouring nodes' density to determine the nodes awaken period can be beneficial to the transmission process. Fig. 13 shows a disparity in the nodes' awaken duration range. For an average density of 60 neighbours, the node awaken duration ranges between 30 000 fs and 100 000 fs. For a density of 100 neighbours,

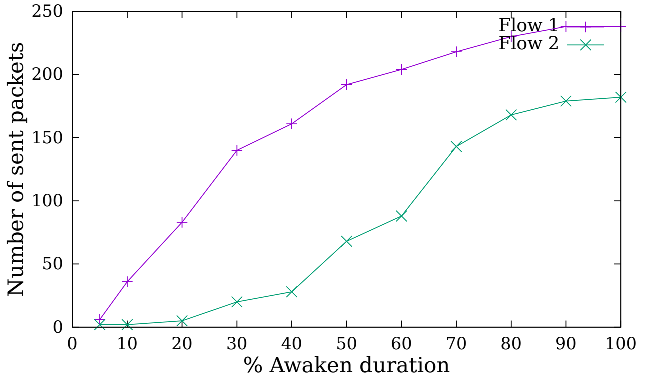


Fig. 12 – The number of packets sent varies with node awaken duration.

the node awaken duration range starts at 50 000 fs. As the average of neighbours' nodes increases, the awaken duration range shrinks and goes close to 100 000 fs.

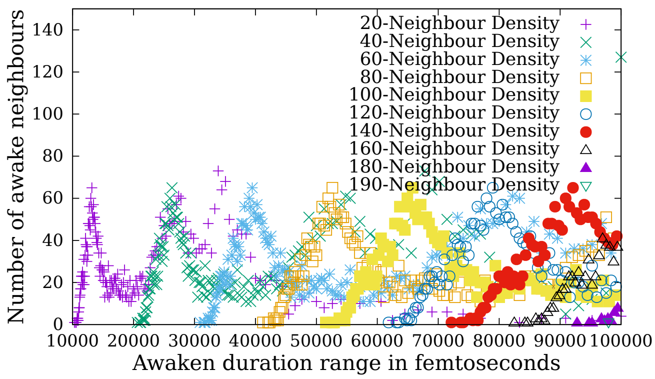


Fig. 13 – Number of awaken nodes in a duration range for different average densities.

Packet reachability can be significantly improved depending on the average density. For example, when specifying awakenNodes=20 to BitSimulator, each node executes a node density estimation algorithm and computes the awake interval according to a simple formula. Starting from 20 as average density, packet reception starts increasing (Fig. 14). An average density of 50 neighbours is sufficient for packets to reach their destinations for flow 1, whilst 60 neighbours for flow 2 are needed. Sleep- ing improves network behaviour by limiting the amount of traffic an individual node can see, but provisions have to be made to ensure the destination node is not sleeping when data packets reach it (example flow 2 at 130 neigh- bours nodes, where the curve has lost its stability). *Note that, in this scenario, an average density of 190 means that all nodes are awake all the time.*

When average density is low, most packets are not transmitted (because packet propagation stops at the beginning or in their path to the destination). Fig. 15 shows that for low average densities (e.g. 20, 30), the number of packets sent is quite low (close to 0). Indeed, with the increasing of the neighbours' nodes average density, packets sent will exponentially increases. For 140 neighbours, the number reaches its highest value at 238

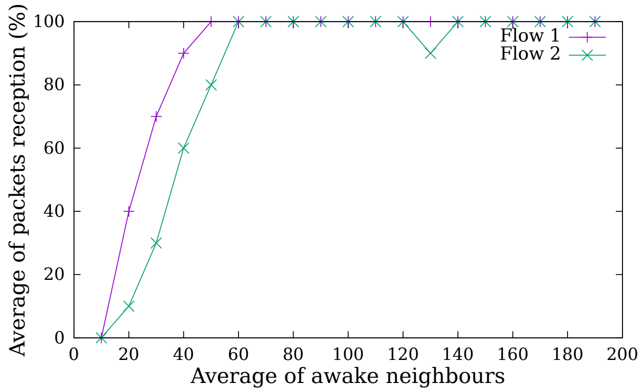


Fig. 14 – Percentage of arrived packets depending on neighbours nodes average, for 2 flows.

for flow 1 and 182 for flow 2. We notice that with 60 neighbours' nodes (*where all packets reach their destination*), the total number of packets sent for flow 1 is 205 and 155 for flow 2.

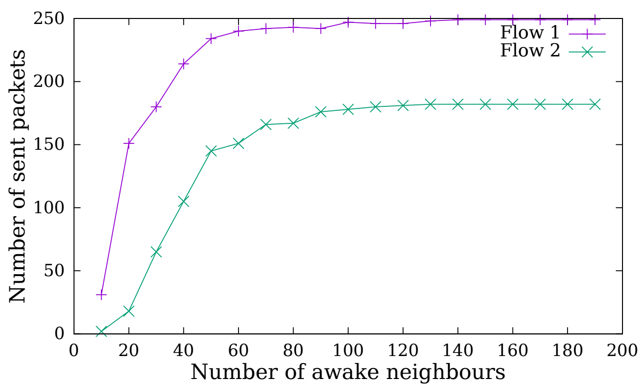


Fig. 15 – Packet transmission cost depending on the average of neighbours' nodes.

5.5 Processing at the destination zone

As explained before, the proposed sleeping mechanism reduces the congestion problem and preserves node resource consumption along the path from the source to the destination. But the very definition of destination may change depending on the application. If the destination is defined as an SLR address, it means that we want the packet to reach this SLR zone and that at least one node in this zone must receive this packet. In that case, the mechanism we proposed is efficient and can be used as is. On the other hand, if we aim to reach a specific node at the destination zone, then more aspects have to be taken into consideration. When a packet arrives at the destination zone, the destination node by chance may be asleep and consequently misses the packet. Different strategies might be used to solve this problem, depending on the node peculiarities, the local density and the application requirement. The algorithm is well explained in Section 4.

Fig. 16 shows the impact of applying our proposed retransmission algorithm to all the nodes at the destination zone. The algorithm clearly enhances the average of packet reception with the flow 1-patch compared to flow1-nopatch (without the algorithm). A value of 40 neighbours nodes ensures an average of 100% of packet reception by the destination node. Indeed, the positive impact is also clearly shown for flow 2-patch at the 130 of awake neighbours, comparing to the flow 2-nopatch, where the destination node missed the packet because it was asleep. The proposed algorithm shows its effect in increasing node chances in receiving the packet if it was asleep at the moment the packet reaches its zone (Fig. 17).

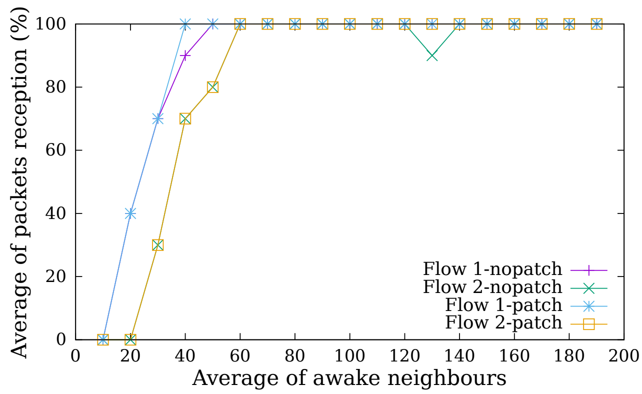


Fig. 16 – The impact of the retransmission algorithm.

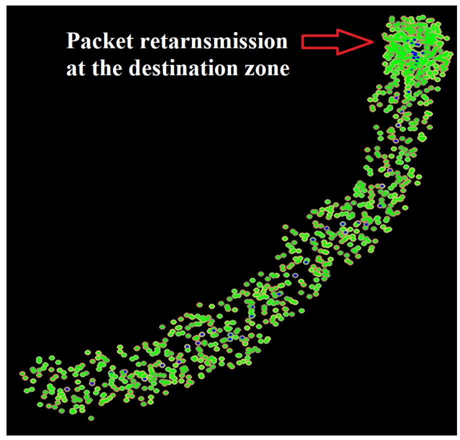


Fig. 17 – The retransmission algorithm applied at the destination zone increases the number of exchanged packets in that zone.

5.6 Average density vs awaken percentage

The node awaken interval, considered the main difference between the two ways of applying a sleeping mechanism. The DEDeN algorithm allows a node to estimate the number of its neighbours. Based on this value, nodes will set their awaken interval (*awaken duration*) (Fig. 18). However, fixing an awaken percentage (e.g. 80%), means that all the nodes in the network will be awake for the same duration (e.g. 80 000 fs).

The longer the node sleeps, the lower the consumed resources. The difference in nodes awaken duration is considered a special peculiarity of the average density (awakenNodes). An average density of 60 neighbours shows an awaken duration distribution (*less resource con-*

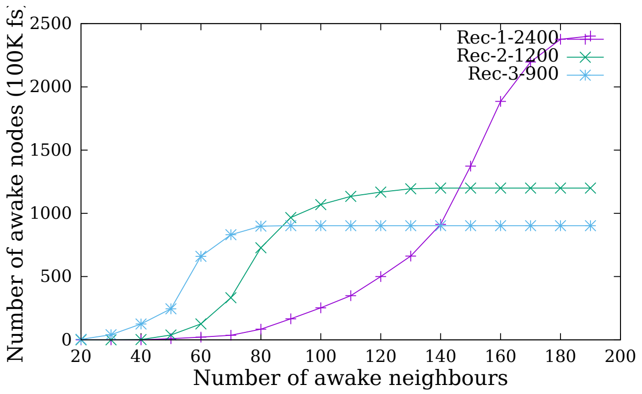


Fig. 18 – Number of full-awake nodes over different average density values.

sumption) ranging between 20 000 and 100 000 fs (*different awaken duration for each node*). Furthermore, for an awaken percentage of 90%, all the nodes in the network (4500 nodes) will be awake along this duration (Fig. 19).

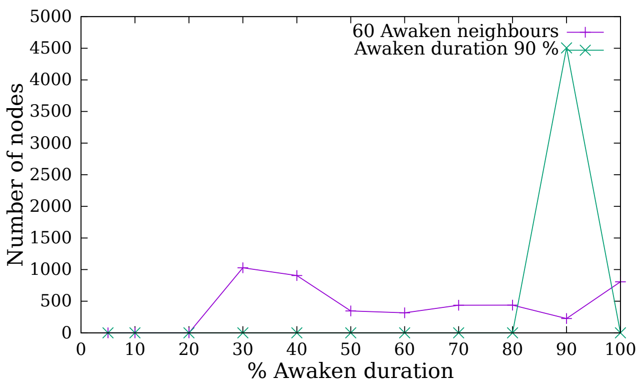


Fig. 19 – Number of nodes depending on an awaken duration of 90% and 60 neighbour nodes.

The total number of sent packets is a metric we rely on to compare between the methods used in applying the sleep mechanism (awakenDuration vs awakenNodes). Based on the results in the previous sections, setting the node awaken duration based on node neighbour density (using DEDeN) will help in decreasing the number of packets sent by 14% when compared to a fixed awaken percentage (awakenDuration).
To conclude, in heterogeneous networks, the SLR routing protocol is acting well when crossing areas of several densities. It smoothly deals with nodes of different awaken durations, and packets are successfully delivered to their intended destination. Applying DEDeN allows nodes to have different awaken durations compared to a static awakenDuration. This leads to a decrease in packet transmission cost, and preservation of node re- sources (CPU, energy, memory), therefore an increase in network lifetime. Moreover, applying the retransmission algorithm at the destination zone increases the chances for the destination node to receive its intended packet. Therefore, this algorithm helps to increase the reliability of data arrival to the destination node, which is an important parameter in some applications.

6. CONCLUSION AND FUTURE WORK

This paper presents the effect of the sleeping mechanism in a heterogeneous nano-network. The new idea embeds a density estimator algorithm (DEDeN) to automatically tune the node awaken interval (awakenDuration). The estimator algorithm can be executed in different times (network deployment, or when a node needs to estimate its number of neighbours). It shows its usefulness when used in conjunction with the sleeping mechanism. It becomes clear that relying on average neighbour node density has a positive influence in preserving node resources (CPU, energy, memory ...) and decreases the number of sent packets compared to results based on static percentage of awaken duration.
Considerations have also been taken for the specific case of the destination zone (packet loss by the destination node because its sleep may coincide with the packet's arrival at the destination zone). An algorithm has been proposed in case the application needs data packets to reach a specific node in the destination zone.
As future work, we would like to consider different algorithms and variations to handle what happens at the destination zone. Especially, a destination could be expressed as "any node in the destination area that provides a given service" instead of all nodes in the zone or a single, specific node.

7. ACKNOWLEDGMENTS

Ali Medlej has a grant from Islamic Center Association for Guidance and Higher Education (Lebanon). This work has also been funded by Pays de Montbéliard Agglomération (France).

REFERENCES

[1] Xin-Wei Yao and Ye-Chen-Ge Wu and Wei Huang, "Routing techniques in wireless nanonetworks: A survey", in Nano Communication Networks, pp. 1–13, Sep, 2019, Elsevier.

[2] Büther, Florian and Lau, Florian-Lennert and Stelzner, Marc and Ebers, Sebastian, "A formal definition for nanorobots and nanonetworks", in Internet of Things, Smart Spaces, and Next Generation Networks and Systems, pp. 214–226, Aug, 2017, Springer.

[3] Akyildiz, Ian F and Jornet, Josep Miquel, "The Internet of nano-things", in IEEE Wireless Communications, pp. 58–63, Dec, 2010.

[4] Jornet, Josep Miquel and Akyildiz, Ian F, "Femtosecond-long pulse-based modulation for terahertz band communication in nanonetworks", in IEEE Transactions on Communications, pp. 1742–1754, Apr, 2014.

[5] Tsioliaridou, Ageliki and Liaskos, Christos and Dedu, Eugen and Ioannidis, Sotiris, "Packet routing in 3D

nanonetworks: A lightweight, linear-path scheme", in Nano Communication Networks, pp. 63–71, Jun, 2017, Elsevier.

[6] Akyildiz, Ian F and Jornet, Josep Miquel, "Electromagnetic wireless nanosensor networks", in Nano Communication Networks, pp. 3–19, March, 2010, Elsevier.

[7] N. Ali and M. Abu-Elkheir, "Internet of nano-things healthcare applications: Requirements, opportunities, and challenges", in IEEE 11th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), Los Alamitos, CA, USA, pp. 9–14, Oct, 2015.

[8] Maksimovi , Mirjana, "The roles of Nanotechnology and Internet of Nano things in healthcare transformation", in TecnoLógicas, pp. 139–153, Dec, 2017.

[9] Seyedi, MirHojjat and Kibret, Behailu and Lai, Daniel TH and Faulkner, Micheal, "A survey on intrabody communications for body area network applications", in IEEE Transactions on Biomedical Engineering, pp. 2067–2079, Aug, 2013.

[10] Lee, Won-Suk and Alchanatis, Victor and Yang, Chenghai and Hirafuji, Masayuki and Moshou, D and Li, Changying, "Sensing technologies for precision specialty crop production", in Computers and electronics in agriculture, pp. 2–33, Oct, 2010, Elsevier.

[11] Giraldo, Juan Pablo and Wu, Honghong and Newkirk, Gregory Michael and Kruss, Sebastian, "Nanobiotechnology approaches for engineering smart plant sensors", in Nature nanotechnology, pp. 541–553, Jun, 2019, Nature Publishing Group.

[12] Wegner, Theodore H and Winandy, Jerrold E and Ritter, Michael A, "Nanotechnology opportunities in residential and non-residential construction", in NICOM 2: 2nd International Symposium on Nanotechnology in Construction, pp. 339–347, Nov, 2006, RILEM Publications SARL.

[13] Zheng, Wei and Shih, Hui-Ru and Lozano, Karen and Mo, Yi-Lung, "Impact of nanotechnology on future civil engineering practice and its reflection in current civil engineering education", in Journal of Professional Issues in Engineering Education and Practice, pp. 162–173, Jul, 2011, American Society of Civil Engineers.

[14] Ali Medlej and Kamal Beydoun and Eugen Dedu and Dominique Dhoutaut, "Scaling up Routing in Nanonetworks with Asynchronous Node Sleeping", in 28th International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Hvar, Croatia, pp. 177–182, Sep, 2020, IEEE.

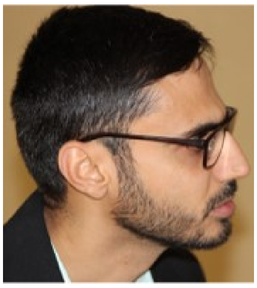
[15] Arrabal, Thierry and Dhoutaut, Dominique and Dedu, Eugen, "Efficient density estimation algorithm for ultra dense wireless networks", in 27th International Conference on Computer Communication and Networks (ICCCN), pp. 1–9, Jul, 2018, IEEE.

[16] Dominique Dhoutaut and Thierry Arrabal and Eugen Dedu, "BitSimulator, an electromagnetic nanonetworks simulator", in 5th ACM/IEEE International Conference on Nanoscale Computing and Communication (NanoCom), New York, NY, USA, pp. 1–6, Sep, 2018, Association for Computing Machinery.

AUTHORS

A. Medlej holds a degree in computer and telecommunications engineering and a specialized master's in telecom, network and security, option: cybersecurity engineering in (2018) from the Lebanese University – Faculty of Engineering and Saint Joseph University of Beirut-Beirut Higher School of Engineering. He starts a co-direction PhD between Franche-Comté University, SPIM Doctoral School, FEMTO-ST lab - Nano-Network Research Team, Montbéliard, France and the Lebanese University, EDST Doctoral School, Faculty of Sciences I, L'ARICoD lab, Hadath, Lebanon, entitled « Congestion control in dense nano-networks» under the direction of Doctor Eugen Dedu (Franche Comte) and the supervision of Doctor Kamal Beydoun (Lebanese University) and Doctor Dominique Dhoutaut (Franche Comte).

E. Dedu has been an associate professor at the University of Franche-Comté (UFC) in Montbéliard, France since 2003. He is a member of the Department of Computer Science and Complex Systems (DISC) of the FEMTO-ST laboratory. He received his Ph.D. in Computer Science at University of Versailles St Quentin, France, in 2003. He received an engineering degree from University Politehnica in Bucarest, Romania, in 1998. He co-authored around 50 articles in international journals and conferences. His current research focuses on wireless dense networks and wireless nanonetworks.



K. Beydoun is an associate professor at the Lebanese University (LU), Beirut, Lebanon, and a member of the Faculty of Sciences since 2011. He obtained his undergraduate degree at the Lebanese University (2001) in computer science. Beydoun received his Ph.D. in computer net- working at Franche-Comte University, Besancon, France under the supervision of Professor Hervé Guyennet. A fellow researcher in LARICoD Laboratory at the Lebanese University collaborating with international research labs and universities, his current research focuses on wireless dense networks and wireless nanonetworks. His work is situated in the crossdisciplinary study of computer science, particularly Wireless Sensor Networks(WSN) and Intrusion Detection System(IDS). Kamal teaches several undergraduate courses in software engineering and programming languages.

D. Dhoutaut is an associate professor at the University of Franche-Comté (UFC) in Montbéliard, France. He is a member of the Department of Computer Science and Complex Systems (DISC) of the FEMTO-ST laboratory. After a PhD on mobile ad hoc networks and IEEE 802.11 at INSA Lyon and also consecutive work on Vehicular Ad Hoc Networks, his research interests now reside in nano-networks. He develops protocols at various layers and also specializes in evaluation tools, such as complex networks simulations software.

HIERARCHICAL BEAM ALIGNMENT IN SU-MIMO TERAHERTZ COMMUNICATIONS

Yifei Wu[1], Johannes Koch[2], Martin Vossiek[2], and Wolfgang Gerstacker[1]
[1]Institute for Digital Communications, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany,
[2]Institute of Microwaves and Photonics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

NOTE: Corresponding author: Yifei Wu, yifei.wu@fau.de

Abstract – *Single-Carrier Frequency Division Multiple Access (SC-FDMA) is a promising technique for high data rate indoor Terahertz (THz) communications in future beyond 5G systems. In an indoor propagation scenario, the Line-Of-Sight (LOS) component may be blocked by the obstacles. Thus, efficient THz SC-FDMA communications require a fast and reliable Beam Alignment (BA) method for both LOS and Non-Line-Of-Sight (NLOS) scenarios. In this paper, we first adopt the hierarchical discrete Fourier transform codebook for LOS BA, and introduce the hierarchical k-means codebook for NLOS BA to improve the beamforming gain. Simulation results illustrate that the hierarchical DFT codebook and the hierarchical k-means codebook can achieve the beamforming gain close to that of the maximum ratio transmission in LOS and NLOS cases, respectively. Based on these two codebooks, we propose a Multi-Armed Bandit (MAB) algorithm named Hierarchical Beam Alignment (HBA) for single-user SC-FDMA THz systems to reduce the BA latency. HBA utilizes a hierarchical structure in the adopted codebook and prior knowledge regarding the noise power to speed up the BA process. Both theoretical analysis and simulation results indicate that the proposed BA method converges to the optimal beam with high probability for both the hierarchical DFT codebook and the hierarchical k-means codebook in the LOS and NLOS scenarios, respectively. The latency introduced by HBA is significantly lower when compared to an exhaustive search method and other MAB-based methods.*

Keywords – Beam alignment, codebook design, SC-FDMA, SU-MIMO, terahertz communications

1. INTRODUCTION

Terahertz (THz) communications are a key technology for the future wireless communications (beyond 5G) owing to its ample frequency spectrum resource between 0.1 and 10 THz promising a much higher capacity than mmWave communications [1], [2]. Major breakthroughs in hardware and theory have been achieved for the efficient realization of THz range transmission [3], [1]. Besides the ultra-high bandwidth, THz wireless technology has the advantage that it could be deployed much faster and more efficiently than optical fiber systems, especially in a high-density urban environment [4]. Moreover, THz communications are highly suited to the indoor environ- ment since THz communications systems can utilize the Non-Line-Of-Sight (NLOS) Multipath Components (MPCs) to enhance the link quality in indoor application [5], [6]. However, NLOS MPCs result in a highly frequency-selective channel, which requires the transmission system to deal with the Inter-Symbol Interference (ISI) effect. The Single-Carrier Frequency Division Multiple Access (SC-FDMA) transmission approach provides a solution to conquer the high frequency selectivity of the channel. Compared to Orthogonal Frequency Division Multiplexing (OFDM), SC-FDMA utilizes a Discrete Fourier Transform (DFT) pre-coding to reduce Peak-to-Average Power Ratio (PAPR) [7]. Due to the high carrier frequency in Time-Division Duplex (TDD) massive Multiple-InputMultiple-Output (MIMO) THz communications, the power amplifier efficiency in the THz band is degraded significantly. Hence, the benefits of a low PAPR of SC-FDMA are emphasized for THz communications. Moreover, THz signals suffer from severe path loss caused by high carrier frequency around hundreds of GHz, which limits THz communications in an indoor transmission scenario [8], [6]. To overcome the high path loss, directional beamforming with massive Transmitter (Tx) and Receiver (Rx) antenna arrays is regarded as a reasonable solution. Because of the short wavelength in THz bands, and the progress of antenna technology, the massive antenna arrays in principle can be packed into a small area, which enables a large beamforming gain at both Tx and Rx. Besides the massive antenna array, the Tx and Rx beams at the base station and user terminal must be formed accurately to achieve the beamforming gain. However, the design of beamforming codes is usually based on perfect Channel State Information (CSI), which is difficult to acquire in the THz case especially at the Tx side due to the large-scale antenna array and the small Signal-to-Noise Ratio (SNR) before beamforming.

One way to circumvent the CSI requirement is to employ a Beam Alignment (BA) scheme. BA is a process to find the optimal transmit-receive beam pair from predefined codebooks to maximize the receive signal strength. The beam alignment problem has been widely studied in mmWave communications. The authors in [9] advocate decoupling the BA process in mmWave transmission into two steps to reduce the BA latency caused by

[0]This work has been supported by the German Research Foundation (DFG) under grant GE 861/9-1.

a large codebook. First, the receiver beam is set to a quasi-omnidirectional beam and the transmitter scans the beam codebook exhaustively for the best transmit beam. Then, the receiver searches for the best beam with quasi-omnidirectional transmit beam. To further accelerate the BA process, we replace the BA step at the receiver by the frequency-domain Minimum Mean-Squared Error (MMSE) equalizer design according to [10] exploiting the high SNR at the receiver after the transmit BA for channel estimation in [11]. Still, the exhaustive search at Tx may take up to several seconds with a massive transmit antenna array.

To reduce the duration of the BA process at the Tx, Multi-Armed Bandit (MAB) theory has been advocated in many search strategies for the mmWave BA problem [12]. MAB focuses on choosing the best arm with maximal expected gain among multiple actions when each choice's gain is known after each selection. Based on the MAB-BA problem framework, many variants of MAB solutions have been adapted to the mmWave communication system. In [13], the authors proposed a celebrated upper confidence bound beam selection scheme in traditional MIMO systems. In [14], the authors introduced a beam alignment algorithm employing contextual information based on a structured MAB framework. The algorithm in [15] utilizes latent probability structure information from the varying transmission environments and promises better performance in fast varying mmWave channels. Another work proposed a distributed BA search method based on adversarial bandit theory, where Tx and Rx choose their beamforming and combining vector independently of each other [16]. Nevertheless, these works do not provide a method with a short training time and accurate beam alignment, especially in multipath channels. Hence, the authors in [9] proposed a Hierarchical Beam Alignment (HBA) algorithm based on a stochastic MAB scheme, leveraging the inherent correlation structure among beams. HBA can identify the optimal beam accurately with a short exploration time, which fulfills the fast BA requirement in narrowband MIMO LOS mmWave communications. However, no specific algorithm is proposed yet for broadband MIMO THz communications. Moreover, if the LOS component of propagation is blocked, the communications quality cannot be guaranteed by the aforementioned algorithms.

In this paper, we consider the design of HBA algorithm based on [9]. Given a hierarchical codebook, HBA selects the beams to maximize the cumulative receive power within a certain period in a hierarchical manner utilizing the hierarchical structure of the codebook. Both theoretical analysis and extensive simulation results demonstrate that HBA can determine the optimal beam with high probability and short latency. Moreover, the hierarchical codebook design for indoor THz propagation is still an open problem. We adopt the hierarchical DFT codebook proposed in [9] to the LOS THz scenario. Furthermore, we propose a data-driven method according to [17] to design a hierarchical codebook utilizing hierarchical k-means clustering for the NLOS THz scenario.

This paper is organized as follows. The system model, channel model and beamforming framework are described in Section 2. In Section 3, the codebook design problems in both LOS and NLOS THz propagation are formulated. The hierarchical DFT codebook is proven to be a local optimum for the LOS codebook problem and is selected as the codebook for LOS HBA. A data-driven codebook design algorithm based on hierarchical k-means clustering is proposed for the NLOS HBA. The HBA algorithm for the SC-FDMA system operating in both LOS and NLOS scenarios is discussed in Section 4. Numerical results are provided in Section 5, followed by conclusions in Section 6.

Notation: Bold lowercase letters (e.g., $\mathbf{a}$) and uppercase letters (e.g., $\mathbf{A}$) represent column vectors and matrices, respectively. $A_{i,j}$ represents the (i,j)-th entry of $\mathbf{A}$. a_k and $\mathbf{a}_k$ stand for the kth entry of $\mathbf{a}$ and the kth column of $\mathbf{A}$, respectively. $(\cdot)^T$, $(\cdot)^*$, $(\cdot)^H$ denote the transpose, conjugate and Hermitian transpose of a vector or matrix, respectively. All complex-valued gradients follow the definition according to [18].

2. SYSTEM MODEL

2.1 System model

The block diagram of the SU MIMO SC-FDMA transmission is shown in Fig. 1.

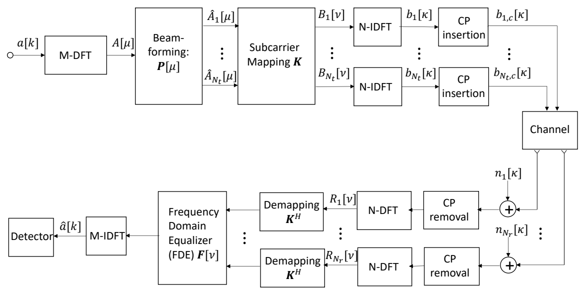


Fig. 1 – Structure of SU-MIMO SC-FDMA transmission system.

Here, the base station and the user terminal are equipped with N_t and N_r antennas, respectively, where N_t is set to be a power of two, e.g., $N_t = 4, 8$, or 16 .Let $a_s[k], k \in \{0, 1, \cdots, M-1\}$ be the transmit data sequence with variance σ_a^2 corresponding to the symbol vector $\mathbf{a}_s$ for transmission in the SC-FDMA system. For simplicity, only a single block is considered in the following. An M-point DFT precoding is applied to symbol block $\mathbf{a}_s$ to transform $\mathbf{a}_s$ to an $M \times 1$ vector $\mathbf{A} = \mathbf{W}_M \mathbf{a}_s$ in the frequency domain with $\mathbf{A} = [A[0], A[1], \cdots, A[M-1]]^T$. Afterwards, an $N_t \times 1$ digital precoding vector $\mathbf{w}[\mu]$ is employed for each frequency domain symbol $A[\mu], 0 \leq \mu \leq M-1$ to form vectors of size $N_t \times 1$ via

$$\hat{\mathbf{A}}[\mu] = \mathbf{w}[\mu] A[\mu] \tag{1}$$

as shown in Fig. 1. Let $\hat{\mathbf{A}}_i = [\hat{A}_i[0], \hat{A}_i[1], \cdots, \hat{A}_i[M-1]]^T$ denote the $M \times 1$ pre-coded frequency domain vector correponding to the ith transmit antenna. N_t vectors $\hat{\mathbf{A}}_i, i \in \{1, 2, \cdots, N_t\}$ are mapped to N subcarriers , $N \geq M$, by the assignment matrix $\mathbf{K}$, leading to frequency domain vectors $\mathbf{B}_i$ with size N which are represented as $\mathbf{B}_i = \mathbf{K}\hat{\mathbf{A}}_i$. We assume the SC-FDMA subcarriers are assigned in a localized mode in this work, where symbols in $\hat{\mathbf{A}}_i$ are mapped to M consecutive subcarriers starting from the ν_0th subcarrier, i.e., $\mathbf{K} = [\mathbf{0}^T_{\nu_0 \times M}\ \mathbf{I}^T_M\ \mathbf{0}^T_{(N-M-\nu_0)\times M}]^T$. After performing the N-point inverse DFT operation on $\mathbf{B}_i$ and adding a Cyclic Prefix (CP) with a length of L_c, the time domain data symbols are transmitted.

At the receiver, first the CP is removed. Subsequently, the time-domain symbols are transformed to the frequency domain by an N-point DFT. The following matrix-vector model holds for the νth subcarrier:

$$\mathbf{R}[\nu] = \mathbf{H}[\nu]\mathbf{B}[\nu] + \mathbf{N}[\nu], \tag{2}$$

where $\mathbf{R}[\nu]$, $\mathbf{H}[\nu]$ and $\mathbf{N}[\nu]$ denote the received signal vector with size $N_r \times 1$, the $N_r \times N_t$ MIMO channel frequency response matrix and the noise vector at the νth subcarrier, where $\mathbf{N}[\nu] \sim \mathcal{N}(\mathbf{0}, \sigma_n^2 \mathbf{I}_{N_T})$, respectively. Vector $\mathbf{B}[\nu] = [B_1[\nu], B_2[\nu], \cdots, B_{N_t}[\nu]]^T$ denotes the $N_t \times 1$ transmit signal vector at the νth subcarrier. Since in the BA phase, the CSI is not known at Rx, the receiver is set to be quasi-omnidirectional. The average power $\bar{p}[\nu]$ of the received vector $\mathbf{R}[\nu]$ is given by

$$\begin{aligned}\bar{p}[\nu] &= \mathbb{E}\{\mathbf{R}^H[\nu]\mathbf{R}[\nu]\} \\ &= \mathbf{w}^H[\nu-\nu_0]\mathbf{H}^H[\nu]\mathbf{H}[\nu]\mathbf{w}[\nu-\nu_0]\sigma_a^2 + N_r \times \sigma_n^2.\end{aligned} \tag{3}$$

The objective of the developed SC-FDMA beamforming is to maximize the average receive signal power $\bar{p}[\nu]$ for each subcarrier under a transmit power constraint. After the BA phase, the CSI can be acquired by the channel estimation approach proposed in [11] due to the large receive signal power after BA. Therefore, we focus on the BA task at the transmitter in this work.

2.2 Channel model

In this work, the indoor transmission scenario in Fig. 2 is considered. The corresponding room has dimensions 5m×5m×3m. The transmitter is fixed at the center of the room ceilling in order to cover the entire room, i.e., at position $(2.5\text{m}, 2.5\text{m}, 3\text{m})$. Meanwhile, the receiver is placed uniformly within the given room. Representing a cellphone or a laptop, the receiver is positioned at a height of 1.5 m. Both transmitter and receiver are equipped with a Uniform Linear Array(ULA) along the y-axis with antenna element spacing r. Indoor THz channel models have been developed in [8], [6],[19],[20]. In this paper, we mainly adopt the deterministic channel model based on the ray-tracing technique according to [6] and [21], since it describes the THz channel more accurately compared to the statistical channel models in [20] and [19].

Based on the sparsity of the given indoor scenario, the received Multipath Components (MPCs) typically comprise a Line-Of-Sight (LOS) component and several Non-Line-Of-Sight (NLOS) components caused by reflection and scattering. Different MPCs have different Angles of Departure (AoDs) and Angles of Arrival (AoAs).The superposition of LOS channel and NLOS channel including transmit and receive filtering results in the discrete-frequency channel matrix $\mathbf{H}[\nu]$ for the νth subcarrier. According to the channel modeling in [6], [22], and [23], the MIMO channel frequency response for the LOS component depends on the subcarrier frequency f_ν and the distance between transmitter and receiver d, and is denoted by $\mathbf{H}_{LOS}(f_\nu, d)$. The MIMO channel frequency response for the NLOS components is a function of the subcarrier frequency f_ν and the geometry information of the propagation scenario collected in vector ζ, and is denoted by $\mathbf{H}_{NLOS}(f_\nu, \zeta)$. Thus, the MIMO channel frequency response for the LOS component $\mathbf{H}_{LOS}(f_\nu, d)$ and the NLOS components $\mathbf{H}_{NLOS}(f_\nu, \zeta)$, respectively, for the νth subcarrier with frequency f_ν are given by

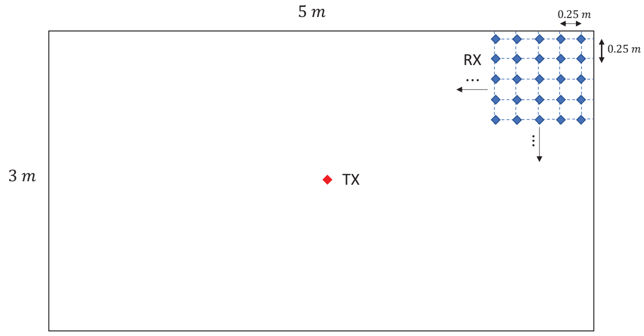


Fig. 2 – Indoor propagation scenario.

$$\begin{aligned}\mathbf{H}_{LOS}(f_\nu, d) &= \alpha_{LOS}(f_\nu, d)\, G_t\, G_r \\ &\quad \mathbf{a}(N_r, \theta_{LOS})\, \mathbf{a}(N_t, \phi_{LOS})^T \\ \mathbf{H}_{NLOS}(f_\nu, \zeta) &= \sum_{i=1}^{n_{NLOS}} \alpha_i(f_\nu, \zeta_i)\, G_t\, G_r \\ &\quad \times \mathbf{a}(N_r, \theta_i(\zeta_i))\, \mathbf{a}(N_t, \phi_i(\zeta_i))^T.\end{aligned} \tag{4}$$

Here, ϕ_{LOS}, $\phi_i(\zeta_i)$, θ_{LOS} and $\theta_i(\zeta_i)$ are the AoD of LOS component, AoDs of the corresponding NLOS components, AoA of LOS component and AoAs of the corresponding NLOS components, respectively. Since the reflection and scattering behaviors for all subcarriers are assumed to be the same, the AoAs and AoDs are modeled as constant over frequency. Their values can be computed accurately by the adopted ray-tracing algorithm based on the geometry information $\zeta_i, i \in \{1, 2, \cdots, n_{NLOS}\}$ of the corresponding scattering scenario. Here, ζ_i contains the distance between transmitter and scattering point $d_{i,1}$, the distance between scattering point and receiver $d_{i,2}$, the incident angle $\theta_{i,1}$, reflected angle $\theta_{i,2}$ and the scattering direction $\theta_{i,3}$. More details on the definition of the mentioned parameters and AoA/AoD calculation can be found in [6]. G_t and G_r are the transmit and receive antenna gains, respectively, and $\mathbf{a}(N_t, \theta)$ and $\mathbf{a}(N_r, \phi)$ denote the transmit and receive array steering vectors, which can be expressed as

$$\mathbf{a}(N_a,\phi) = \left[1,\cdots,e^{j\frac{2\pi f_\nu r}{c}n\sin\phi},\cdots,e^{j\frac{2\pi f_\nu r}{c}(N_a-1)\sin\phi}\right]^T, \quad (5)$$

where n is the antenna element index with $0 \leq n \leq N_a - 1$. N_a denotes the number of antennas, r is the antenna spacing and c is the speed of light. In an indoor THz channel, the LOS path gain $|\alpha_{LOS}(f_\nu,d)|^2$ depends on the spreading loss $L_{spread}(f_\nu,d)$ and absorption loss $L_{abs}(f_\nu,d)$,

$$\begin{aligned} |\alpha_{LOS}(f_\nu,d)|^2 &= L_{spread}(f_\nu,d)L_{abs}(f_\nu,d), \\ L_{spread}(f_\nu,d) &= \left(\frac{c}{2\pi f_\nu d}\right)^2, \\ L_{abs}(f_\nu,d) &= e^{-k_{abs}(f_\nu d)}. \end{aligned} \quad (6)$$

In (6), d stands for the distance between transmitter and receiver, and $k_{abs}(f_\nu,d)$ denotes the frequency-dependent medium absorption coefficient. The NLOS components depend additionally on the reflection coefficient of the correponding surface. The squared magnitude of the ith NLOS coefficient assuming first-order reflection of the reflecting surface is given by

$$\begin{aligned} |\alpha_i(f_\nu,\zeta_i)|^2 = \Gamma_i^2(f_\nu,\zeta_i)\left(\frac{c}{2\pi f_\nu(d_{i,1}+d_{i,2})}\right)^2 \\ \times e^{-k_{abs}(f_\nu(d_{i,1}+d_{i,2}))}, \end{aligned} \quad (7)$$

where $\Gamma_i(f_\nu,\zeta_i)$ is the product of the Fresnel reflection coefficient and the scattering coefficient and can be calculated as given in [6]. Due to the high reflection loss in the THz band, the restriction to first-order reflection MPCs is justified.

In [22], the authors advocated the extension of a statistical channel model to the THz case in order to characterize the reflection behavior of the furniture. As discussed in [8], [24] and [22], modeling the reflection components due to furniture by the ray-tracing technique is not feasible due to an extremely high computational load. Thus, we adopt the hybrid channel modeling idea in [22] that the LOS component and the reflection components due to the walls and ceiling are generated by the ray-tracing technique and the reflection components due to the furniture are generated via the Saleh-Valenzuela (S-V) channel model. We adopt the parameters' setting of the S-V channel model in [22]. The details of this hybrid channel modeling can be found in [22].

2.3 Beamforming framework

From a system level, the beamforming at the transmitter is required to maximize the sum receive power over the transmission band. If $\mathbf{H}[\nu]$ for all $\nu = 1,2,\cdots,N$ is known at the transmitter, the optimal beamforming vector can be computed using the approach provided in [10]. However, in THz communications it is not practical to perform an entry-wise estimation of the channel matrix at the transmitter, which has a large scale due to the large antenna arrays. Thus, selecting the beam code from a predefined codebook is frequently considered in THz communications.

For a LOS scenario in THz band, the channel is dominated by the LOS component assuming its power is significantly larger than that of the other MPCs [8], [6]. In such a case, the beamforming problem is simplified to finding the AoD of the LOS component, and setting the beamforming vector as the steering vector pointing to the AoD of the LOS component. Thus, the codebook is composed only of steering vectors. In addition, in the LOS scenario, the coherence bandwidth is relatively larger due to the dominance of the LOS component, i.e., the LOS channel exhibits a flatter frequency response than that of the NLOS channel. Thus, neighboring subcarriers can employ the same beamforming vector.

However, in an indoor THz scenario, the LOS propagation might be blocked by the presence of obstacles like moving persons, furniture, or many diverse objects [6]. In that case, the codebook cannot be designed as the composition of steering vectors anymore but has to be specifically tailored to the NLOS scenario. Also, the NLOS channel may be frequency-selective due to the absence of the LOS component. Therefore, each subcarrier should be assigned a carefully designed code word.

According to the aforementioned difference between the LOS beamforming and the NLOS beamforming, a prejudgement regarding the propagation scenario has to be conducted before beamforming. Since the LOS component power is nearly 20 dB larger than that of the NLOS components in the considered typical indoor scenario, the channel can be classified by the receive power. For a training block, if the measured receive power is higher than a threshold P_{th}, the channel will be classified to a LOS channel, and the LOS beamforming procedure will be performed in the BA phase. Otherwise, the channel will be considered as a NLOS channel and the procedure for the NLOS beamforming will take place in the following BA phase. The power threshold P_{th} depends on the geometry of the propagation scenario, the humidity of the atmosphere and the material of the reflecting objects. Thus, the power threshold must be designed based on the given propagation scenario and no general rule regarding its selection can be introduced. The adopted power threshold choice will be discussed in detail in Section 3. After the channel assessment, the beamforming vectors for each subcarrier are determined by the corresponding BA procedure, which will be discussed in detail in Section 4.

3. CODEBOOK DESIGN

In this section, we discuss the hierarchical beam codebook used in hierarchical beam alignment. Hierarchical codebooks have been already studied in [25]. The hierarchical codebook is also the core of the hierarchical beam alignment, which helps to improve the efficiency in searching for the best beam code. In this section, we first introduce the definition of the hierarchical codebooks,

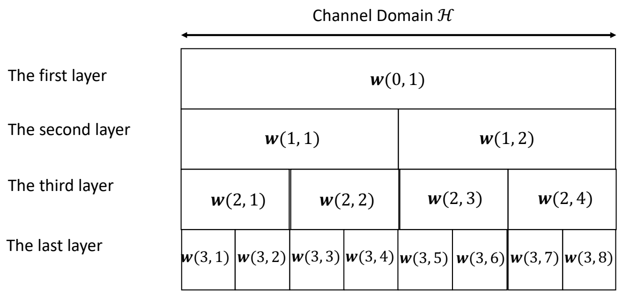


Fig. 3 – The beam coverage of a 3-layer hierarchical codebook.

then discuss the design of the hierarchical codebooks in the LOS scenarios and the NLOS scenarios.

3.1 Definition

Before introducing the definition of a hierarchical codebook, we first define the beam coverage $CV(\mathbf{w}_i)$ of a code word $\mathbf{w}_i$ within a codebook **W**. Let $\mathcal{H}$ denote the set of all possible channel snapshots within the given indoor propagation scenario, which is called the channel domain in the following. The beam coverage $CV(\mathbf{w}_i)$ of a code word **w** in a codebook **W** is a subset of the channel domain $\mathcal{H}$, where the receive power with $CV(\mathbf{w}_i)$ is strongest among all code words in **W**, i.e.,

$$CV(\mathbf{w}_i) = \left\{\mathbf{H} \,\middle|\, \|\mathbf{H}\mathbf{w}_i\|_2 = \max\nolimits_{\mathbf{w}_j \in \mathbf{W}} \|\mathbf{H}\mathbf{w}_j\|_2, \mathbf{H} \in \mathcal{H}\right\}. \tag{8}$$

Note that for each channel realization **H** in $\mathcal{H}$, there is always a code word in **W** aligned to **H**. In other words, the union of the beam coverages of all code words composes the channel domain $\mathcal{H}$, i.e.,

$$\cup_{\mathbf{w}_j \in \mathbf{W}} CV(\mathbf{w}_j) = \mathcal{H}. \tag{9}$$

A hierarchical codebook is a family of I sub-codebooks $\mathbf{W}^i, 1 \leq i \leq I$. In this work, we only consider the binary hierarchical codebook, which means the size of the ith codebook is twice of the size of the $(i-1)$th codebook. The code words in the sub-codebook are layered due to their beam coverage as shown in Fig. 3. Let $\mathbf{w}(i,j), 1 \leq i \leq I, 1 \leq j \leq 2^{i-1}$ denote the jth code word in the ith layer. The beam coverage of all code words in the first $I-1$ layers must satisfy the following criterion:

Criterion 1: The beam coverage $CV(\mathbf{w}(i,j))$ of an arbitary code word $\mathbf{w}(i,j)$ should be the union of those of two code words in the next layer, i.e.,

$$CV(\mathbf{w}(i,j)) = CV(\mathbf{w}(i+1,2j-1)) \cup CV(\mathbf{w}(i+1,2j)), \tag{10}$$

where $0 \leq i \leq I-1$. $\mathbf{w}(i,j)$ is called the parent of $\mathbf{w}(i+1,2j-1)$ and $\mathbf{w}(i+1,2j)$, and $\mathbf{w}(i+1,2j-1)$, $\mathbf{w}(i+1,2j)$ are called the children of $\mathbf{w}(i,j)$.

The above criterion guarantees the codebook is tree-like, which can be used in the hierarchical beam alignment. In the HBA algorithm, which will be introduced in Section 4, the algorithm converges to a beam code in the highest layer with high probability. Thus, the average beamforming gain of a given hierarchical codebook $\mathbf{W} = \{\mathbf{W}^1, \mathbf{W}^2, \cdots, \mathbf{W}^I\}$ is defined as the average receive power $\bar{r}(\mathbf{W})$ of the best beam code in the highest layer, i.e.,

$$\bar{r}(\mathbf{W}) = \int_{\mathbf{H} \in \mathcal{H}} \max\nolimits_{\mathbf{w}(I,j) \in \mathbf{W}^I} \|\mathbf{H}\mathbf{w}(I,j)\|_2^2 f_{\mathbf{H}}(\mathbf{H})\, d\mathbf{H}, \tag{11}$$

where $f_{\mathbf{H}}(\mathbf{H})$ is the Probability Density Function (PDF) of **H**. The aim of the codebook design is to generate a codebook **W** satisfying Criterion 1 meanwhile maximizing the beamforming gain $\bar{r}(\mathbf{W})$ over $\mathcal{H}$. Meanwhile, the transmit power must be limited. Without loss of generality, we assume that the j-th beamforming code word $\mathbf{w}(i,j)$ in the i-th layer has unit norm, i.e., $\mathbf{w}^H(i,j)\mathbf{w}(i,j) = 1$.

The codebook design problem introduced above is formulated on the basis of the codebook design in [25]. However, in [25], mainly the criteria for a hierarchical codebook are proposed and the objective function of the codebook design problem is not explicitly stated. In this work, the two criteria for the hierarchical codebook in [25] are modified to a single criterion by adjusting the definition of the beam coverage. In addition, the beamforming gain of the codes in a hierarchical codebook is utilized in this work as an objective function, which can be regarded as a supplement of the codebook design in [25].

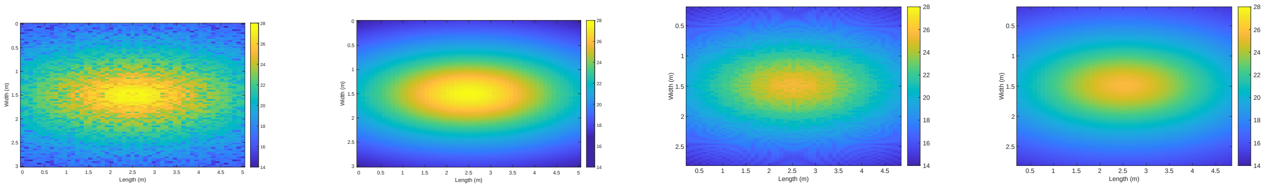
3.2 LoS codebook design

In this subsection, a LoS propagation scenario is considered, where the LoS component is always obtained. Based on the measurement results in [20], the power of the LOS component is nearly 20 dB larger than the power of the NLOS components in the THz band in a typical propagation scenario due to the high reflection and scattering loss. Therefore, the channel in THz LOS scenario is in principle LOS-dominant. In this case, the beam should be focused on the LOS component.

In this paper, the receiver position is assumed to be uniformly distributed within the considered indoor environment in Fig. 2. Thus, the LOS spatial angle $\Phi = \frac{2\pi f_\nu r}{c} \sin\phi_{LOS}$ is considered as uniformly distributed within $(0, 2\pi)$, where r is the antenna spacing. Due to the LOS-dominant channel property in the considered scenario, the beamforming gain of a given hierarchical codebook $\mathbf{W} = \{\mathbf{W}^1, \mathbf{W}^2, \cdots, \mathbf{W}^I\}$ is defined as the the receive power of the best beam code in the highest layer for the LOS component with spatial angle Φ, i.e.,

$$g(\Phi) = \max\nolimits_{\mathbf{w} \in \mathbf{W}^I} \|\mathbf{a}_s^T(N_T, \Phi)\mathbf{w}\|_2^2. \tag{12}$$

where $\mathbf{a}_s(N_T, \Phi) = \left[1, \cdots, e^{j\Phi}, \cdots, e^{j(N_T-1)\Phi}\right]^T$. The codebook design problem is equivalent to maximizing the average beamforming gain over $\Phi \in (0, 2\pi)$. Hence, the



(a) Beamforming gain in dB of the hierarchical DFT codebook in the LOS indoor scenario without furniture.

(b) Beamforming gain in dB of the maximum ratio transmission in the LOS indoor scenario without furniture.

(c) Beamforming gain in dB of the hierarchical DFT codebook in the LOS indoor scenario with furniture.

(d) Beamforming gain in dB of the maximum ratio transmission in the LOS indoor scenario with furniture.

Fig. 4 – Beamforming gain of different codebooks in the given indoor propagation scenario with a carrier frequency of 100 GHz single-frequency transmission.

codebook design problem can be written as

$$\begin{aligned} \max_{\mathbf{W}} \quad & \int_{\Phi=0}^{2\pi} g(\Phi) d\Phi \\ \text{s.t.} \quad & \mathbf{w}^H(i,j)\mathbf{w}(i,j) = 1, 1 \le i \le I, 1 \le j \le 2^{i-1} \\ & CV(\mathbf{w}(i,j)) = CV(\mathbf{w}(i+1,2j-1)) \cup \\ & CV(\mathbf{w}(i+1,2j)),\ 1 \le i \le I-1, 1 \le j \le 2^{i-1}, \end{aligned} \tag{13}$$

where $\mathbf{w}(i,j)$ denotes the jth code word in the ith layer. The above optimization problem is a non-convex problem due to the non-convex constraints. Hence, it is difficult to obtain the globally optimum solution for the codebook $\mathbf{W}$. However, the objective function value depends only on the sub-codebook in the highest layer, i.e., $\mathbf{W}^I = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_{2^{I-1}}^T]$. One greedy approach for the codebook design problem is to choose the sub-codebook in the highest layer $\mathbf{W}^I$ for maximization of the objective function value. Then, the highest sub-codebook $\mathbf{W}^I$ is extended to a hierarchical codebook. The design problem for the highest layer sub-codebook $\mathbf{W}^I$ is given by

$$\begin{aligned} \max_{\mathbf{W}^I} \quad & \int_{\phi} \max_{\mathbf{w} \in \mathbf{W}^I} \|\mathbf{a}^T(N_T,\phi)\mathbf{w}\|_2^2 \, d\phi \\ \text{s.t.} \quad & \mathbf{w}^H(I,j)\mathbf{w}(I,j) = 1, 1 \le j \le 2^{I-1}. \end{aligned} \tag{14}$$

This optimization problem is non-convex since the objective function is a non-concave function. Theorem 1 shows that the DFT codebook will be one of the locally optimum solutions for the above optimization problem if the highest layer sub-codebook size is fixed to N_T. In [9], the unitary DFT matrix $\mathbf{W}_{N_t}$ is chosen as the codebook for the beam alignment, which is given by

$$\mathbf{W}_{N_t} = \frac{1}{\sqrt{N_t}}[\mathbf{a}_s(N_t,\Phi_1), \cdots, \mathbf{a}_s(N_t,\Phi_n), \cdots, \mathbf{a}_s(N_t,\Phi_{N_t})], \tag{15}$$

with $\mathbf{a}_s(N_t,\Phi_n) = \left[1, e^{j\Phi_n}, \cdots, e^{j(N_t-1)\Phi_n}\right]^T$. Here, $\Phi_n = \frac{\pi(2n-1)}{N_t}$, $n \in \{1, \cdots, N_t\}$ denotes the spatial angle of the n-th beam.

Theorem 1: The unitary DFT matrix $\mathbf{F}$ is one of the locally optimum solutions for the above optimization problem if the codebook size is fixed to N_T

Proof. See Appendix A.

Fig. 4 illustrates the beamforming gain for the LOS case in the given indoor scenario. Here, 100×100 sampling points are considered and their beamforming gains are obtained. The beamforming gain definition can be found in (11), which is the objective function of the proposed codebook design. The performance of the resulting codebook can be evaluated by the beamforming gain shown in Fig. 4.

In Fig. 4(a) and 4(b), only the multipath component generated by a ray-tracing algorithm is considered. Fig. 4(a) illustrates the beamforming gain of the DFT codebook in the considered LOS indoor propagation scenario. Compared to the beamforming gain of the Maximum Ratio Transmission (MRT) in Fig. 4(b), which is considered as the upper bound of the beamforming gain, the beamforming gain of the DFT codebook within the entire indoor environment is close to that of MRT. Hence, the hierarchical DFT codebook entails only a slight performance loss with a small codebook size, which can be considered as a welltailored codebook for the LOS scenario.

Fig. 4(c) and Fig. 4(d) illustrate the beamforming gain of the DFT codebook and MRT, respectively, for the LOS case in the given indoor scenario with furniture. The beamforming gains of 100×100 sampling points have been obtained. For one sampling point, 100 statistical channel snapshots have been generated to obtain the average beamforming gain. The difference between the beamforming gain of the DFT codebook and MRT is insignificant, which suggests that the DFT codebook achieves a promising performance even in a complex indoor scenario.

To generate the sub-codebooks in lower layers, the beam coverage of the DFT codebook should be determined. Based on the definition of the beam coverage in Criterion 1, the beam coverage of a beam code $\mathbf{w}(I,j)$ in It h layer is given by

$$CV(\mathbf{w}(I,j)) = \left(\frac{2\pi(j-1)}{N_T}, \frac{2\pi j}{N_T}\right). \tag{16}$$

According to Criterion 1 of the hierarchical codebook, the beam coverage of a beam code $\mathbf{w}(i, j)$, $i = 1, 2, \cdots, I-1$ is determined by

$$CV(\mathbf{w}(i,j)) = \left(\frac{2\pi(j-1)2^{I-i}}{N_T}, \frac{2\pi j 2^{I-i}}{N_T}\right). \tag{17}$$

In the LOS scenario, a code word $\mathbf{w}(i,j)$ is identical to the steering vector located in the center of its corresponding beam coverage $CV(\mathbf{w}(i,j))$, which is given by

$$\mathbf{w}(i,j) = \mathbf{a}(N_T, \Phi_{i,j}), \tag{18}$$

where $\Phi_{i,j} = \frac{\pi(2j-1)2^{I-i}}{N_T}$. *Theorem 1:* The beam coverage $CV(\mathbf{w}(i,j))$ of the j-th code word in the i-th layer can be expressed as

$$CV(\mathbf{w}(i,j)) = \left\{\Phi \,\middle|\, \Phi \in \left[\frac{2\pi(2j-2)}{2^i}, \frac{4j\pi}{2^i}\right]\right\}.$$

Proof. See Appendix B.

3.3 NLOS codebook design

In an indoor scenario, the LOS component might be blocked by obstacles such as persons, furniture, or many other diverse objects. In this case, utilizing the power from NLOS components will be essential for the beamforming performance. Consider a hierarchical codebook $\mathbf{W} = \{\mathbf{W}^1, \mathbf{W}^2, \cdots, \mathbf{W}^I\}$ with I layers. The codebook design problem can be written as

$$\begin{aligned} \max_{\mathbf{W}} \quad & \int_\zeta \max_{\mathbf{w}(I,j)\in\mathbf{W}^I} \|\mathbf{w}(I,j)\mathbf{H}_{NLOS}(f_\nu,\zeta)\|_2^2 \\ & f_\zeta(\zeta)\mathrm{d}\zeta \\ \text{s.t.} \quad & \mathbf{w}^H(i,j)\mathbf{w}(i,j) = 1, \\ & i = 1,2,\cdots,I, j = 1,2,\cdots,2^{i-1} \\ & CV(\mathbf{w}(i,j)) \\ & = CV(\mathbf{w}(i+1,2j-1)) \cup CV(\mathbf{w}(i+1,2j)) \end{aligned} \tag{19}$$

where $f_\zeta(\zeta)$ is the PDF of the geometry information. However, modeling the NLOS components' behavior in the THz band is still an open problem. It is difficult to model the NLOS channel analytically including the reflection and scattering behavior. Faced with this challenge in the NLOS scenario, it is generally intricate to find an analytical method to generate the codebook in the NLOS scenario. Therefore, a data-driven codebook design such as the one proposed in [17] is of great benefit for the considered indoor THz NLOS scenario. In the data-driven approach, the integral over the indoor scenario is approximated by the average over N_p realizations of the indoor channel. In particular, the objective function is defined as

$$f_{NLOS}(\mathbf{W}) = \sum_{i=1}^{N_p} \max_{\mathbf{w}(I,j)\in\mathbf{W}^I} \|\mathbf{H}_{NLOS}^{\nu,i}\mathbf{w}(I,j)\|_2^2, \tag{20}$$

where $\mathbf{H}_{NLOS}^{\nu,i}$ is the NLOS channel frequency response of it h realization at frequency f_ν. The corresponding modified codebook design problem reads

$$\begin{aligned} \max_{\mathbf{W}} \quad & f_{NLOS}(\mathbf{W}) \\ \text{s.t.} \quad & \mathbf{w}^H(i,j)\mathbf{w}(i,j) = 1, \\ & i = 1,2,\cdots,I, j = 1,2,\cdots,2^{i-1} \\ & CV(\mathbf{w}(i,j)) = \\ & CV(\mathbf{w}(i+1,2j-1)) \cup CV(\mathbf{w}(i+1,2j)). \end{aligned} \tag{21}$$

The above optimization problem cannot be solved directly since the second hierarchical structure constraint is non-convex. Hence, before solving the overall hierarchical codebook design problem, a single layer codebook design problem is solved since the objective function depends only on the highest layer sub-codebook. For a single layer codebook $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_J]$, where J is the codebook size, the optimization problem is given by

$$\begin{aligned} \max_{\mathbf{W}} \quad & \sum_{i=1}^{N_p} \max_{\mathbf{w}_j\in\mathbf{W}} \|\mathbf{H}_{NLOS}^{\nu,i}\mathbf{w}_j\|_2^2 \\ \text{s.t.} \quad & \mathbf{w}_j^H\mathbf{w}_j = 1, j = 1,2,\cdots,J \end{aligned} \tag{22}$$

Theorem 3: Let $\hat{\mathbf{H}}_{\nu,i}$ and denote $(\mathbf{H}_{NLOS}^{\nu,i})^H\mathbf{H}_{NLOS}^{\nu,i}$. The distance $D(\mathbf{X},\mathbf{Y})$ between two $N_t \times N_t$ matrices $\mathbf{X}$ and $\mathbf{Y}$ is defined as

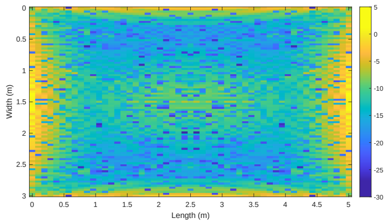
$$D(\mathbf{X},\mathbf{Y}) = \mathrm{tr}((\mathbf{X}-\mathbf{Y})(\mathbf{X}-\mathbf{Y})^H). \tag{23}$$

The optimization problem in *(22)* is equivalent to a distance minimization problem, i.e.,

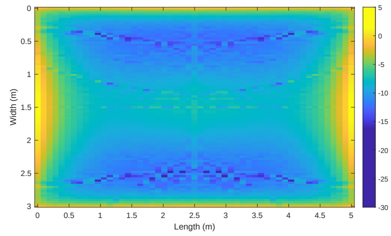
$$\begin{aligned} \min_{\mathbf{W}} \quad & \sum_{i=1}^{N_p} \sum_{\nu=\nu_0}^{\nu_0+M-1} \min_{\mathbf{w}_j\in\mathbf{W}} D(\hat{\mathbf{H}}_{\nu,i}, \mathbf{w}_j\mathbf{w}_j^H) \\ \text{s.t.} \quad & \mathbf{w}_j^H\mathbf{w}_j = 1, j = 1,2,\cdots,J \end{aligned} \tag{24}$$

Proof. See Appendix C.

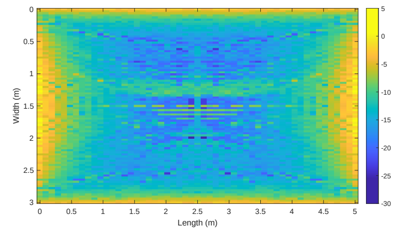
One well-known algorithm to solve the distance minimization problem is k-means clustering, which is an unsupervised machine learning algorithm. Here $\hat{\mathbf{W}}_j, j = 1,2,\cdots,J$ and $\hat{\mathbf{H}}_{\nu,i}, i = 1,2,\cdots,N_p$ are regarded as the clustering centers in k-means clustering and the training data set, respectively. The overall training data set is denoted as $\mathcal{H}_{NLOS}$. The algorithm proceeds by alternating between assignment step and update step as described in the following.



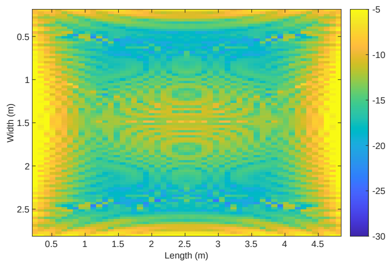
(a) Beamforming gain in dB of the hierarchical DFT codebook in the NLOS case without furniture.



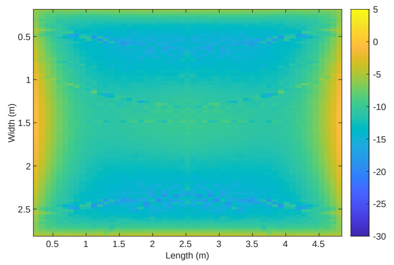
(b) Beamforming gain in dB of the maximum ratio transmission in the NLOS case without furniture.



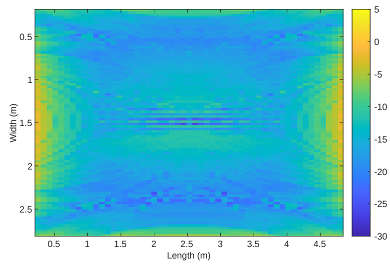
(c) Beamforming gain in dB of the hierarchical k-means codebook in the NLOS case without furniture.



(d) Beamforming gain in dB of the hierarchical DFT codebook in the NLOS case with furniture.



(e) Beamforming gain in dB of the maximum ratio transmission in the NLOS case with furniture.



(f) Beamforming gain in dB of the hierarchical k-means codebook in the NLOS case with furniture.

Fig. 5 – Beamforming gain of different approaches for the given indoor NLOS propagation scenario with carrier frequency of 100 GHz single-frequency transmission.

Algorithm 1 Hierarchical k-means clustering

Input: $\mathcal{H}_{NLOS}, I$
Output: $\mathbf{W} = \{\mathbf{W}^1, \mathbf{W}^2, \cdots, \mathbf{W}^I\}$

Initialization: $\tilde{\mathbf{H}}_1^1 = \sum_{\mathbf{H}_{\nu,i} \in \mathcal{H}_{NLOS}} (\mathbf{H}_{\nu,i})$;
$\mathbf{w}_{1,1} = \text{argmax}_{\mathbf{w}^H\mathbf{w} \leq 1} \mathbf{w}^H \tilde{\mathbf{H}}_1^1 \mathbf{w}$
for all $2 \leq h \leq I$ **do**
  for all $1 \leq j \leq 2^{h-2}$ **do**
    Initialization: generate the initial codebook $\mathbf{W}^{h,j} = \{\mathbf{w}(h, 2j-1), \mathbf{w}(h, 2j)\}$ and $\mathbf{W}_{old}^{h,j} = \mathbf{0}$;
    while $\mathbf{W}^{h,j} \neq \mathbf{W}_{old}^{h,j}$ **do**
      $\mathbf{W}_{old}^{h,j} = \mathbf{W}^{h,j}$;
      $CV(\mathbf{w}(h, 2j-1)) = CV(\mathbf{w}(h, 2j)) = \emptyset$
      $\tilde{\mathbf{H}}_{h,2j-1} = \tilde{\mathbf{H}}_{h,2j} = \mathbf{0}$;
      $\hat{\mathbf{W}}_{h,2j} = \mathbf{w}(h, 2j)\mathbf{w}^H(h, 2j)$
      $\hat{\mathbf{W}}_{h,2j-1} = \mathbf{w}(h, 2j-1)\mathbf{w}^H(h, 2j-1)$
      for all $\mathbf{H}_{\nu,i} \in CV(\mathbf{w}(h-1, j))$ **do**
        if $D(\mathbf{H}_{\nu,i}, \hat{\mathbf{W}}_{h,2j}) \leq D(\mathbf{H}_{\nu,i}, \hat{\mathbf{W}}_{h,2j-1})$ **then**
          $k = 2j$;
        else
          $k = 2j-1$;
        end if
        $CV(\mathbf{w}(h, k)) = CV(\mathbf{w}(h, k)) \cup \mathbf{H}_{\nu,i}$ and $\tilde{\mathbf{H}}_{h,k} = \tilde{\mathbf{H}}_{h,k} + \mathbf{H}_{\nu,i}$;
      end for
      solve $\mathbf{w}(h, k) = \text{argmax}_{\mathbf{w}^H\mathbf{w} \leq 1} \mathbf{w}^H \tilde{\mathbf{H}}_{h,k} \mathbf{w}$ for $k = 2j-1$ and $k = 2j$
    end while
  end for
end for

Assignment step: Assign each training channel $\hat{\mathbf{H}}_i$ to the corresponding clustering center $\hat{\mathbf{W}}_j$, which provides the minimum distance $D(\hat{\mathbf{H}}_i, \hat{\mathbf{W}}_j)$. This means the training channel set is divided into J clusters, i.e., $\mathcal{H}_1, \mathcal{H}_2, \cdots, \mathcal{H}_J$. The jth cluster is expressed mathmatically as

$$\mathcal{H}_j = \{\hat{\mathbf{H}}_{\nu,i} | j = \text{argmin}_{1 \leq j \leq J} D(\hat{\mathbf{H}}_{\nu,i}, \hat{\mathbf{W}}_j)\} \tag{25}$$

Update step: Recalculate centers for the training channels assigned to each cluster. This is done by solving the following optimization problem for $j = 1, 2, \cdots, J$

$$\begin{aligned} \min_{\mathbf{W}} \quad & \sum_{\mathbf{H} \in \mathcal{H}_j} D(\mathbf{H}, \mathbf{W}) \\ \text{s.t.} \quad & \text{tr}(\mathbf{W}) = 1, j = 1, 2, \cdots, J \\ & \text{rank}(\mathbf{W}) = 1. \end{aligned} \tag{26}$$

Theorem 4: The globally optimal solution of $\mathbf{W}$ for *(26)* is given by

$$\mathbf{W} = \mathbf{w}_{max} \mathbf{w}_{max}^H, \tag{27}$$

where $\mathbf{w}_{max}$ is the eigenvector of $\sum_{\mathbf{H} \in \mathcal{H}_j} \mathbf{H}$ corresponding to its largest eigenvalue.

Proof. See Appendix D.

However, the aforementioned approach only generates a single layer codebook, which cannot be adopted for HBA. To guarantee the hierarchical structure of the final resul-ting codebook, one variant of k-means clustering, named hierarchical k-means clustering, is introduced here. The procedure of hierarchical k-means clustering is shown in Algorithm 1. The most important property of hierarchical k-means clustering is that it guarantees the hierarchical structure of the resulting codebook. Moreover, the hierarchical k-means algorithm can speed up k-means-based learning approaches. In the hierarchical k-means algorithm the branching factor is chosen to 2, defining the number of clusters at each layer of the codebook. In the beginning, the beam coverage of $\mathbf{w}(1,1)$ is set to $\mathcal{H}_{NLOS}$. $\mathbf{w}(1,1)$ is selected as the eigenvector belonging to the largest eigenvalue of $\tilde{\mathbf{H}}_1^1 = \sum_{\mathbf{H}_{\nu,i}\in\mathcal{H}_{NLOS}}(\mathbf{H}_{\nu,i})$. Then, $\mathcal{H}_{NLOS}$ is clustered into two sub-clusters $CV(\mathbf{w}(2,1))$ and $CV(\mathbf{w}(2,1))$ by k-means clustering. We recursively further cluster each sub-cluster until we reach the required depth H of the codebook.

Fig. 5 shows the beamforming gain for the NLOS case in the given indoor scenario. Again, in Fig. 5(c), 5(c) and 5(b) only the multipath components generated by the ray-tracing algorithm are considered to obtain the beamforming gain profile. The beamforming gain of the hierarchical k-means codebook generated by the hierarchical k-means clustering is depicted in Fig. 5(c). Compared to the beamforming gain of the hierarchical DFT codebook and that of the MRT in Fig. 5(a) and Fig. 5(b), respectively, the hierarchical k-means codebook achieves a beamforming gain closer to that of the MRT than that of the DFT codebook. Hence, the designed hierachical codebook is a better choice than the DFT codebook in the considered NLOS propagation scenario.

Figures 5(e), 5(a) and 5(f) illustrate the beamforming gain of MRT, the DFT codebook and the hierarchical k-means codebook, respectively, for the NLOS case in the indoor scenario with furniture. Again, 100×100 sampling points have been simulated and 100 statistical channel snapshots have been generated for each sampling point to obtain the average beamforming gain. Based on figures 5(e), 5(a) and 5(f), it can be concluded that in the indoor scenario with furniture, the proposed K-means codebook outperforms the DFT codebook.

From Fig. 5, it can be concluded that the maximal beamforming gain in the NLOS scenario is 5 dB. Meanwhile, the minimal beamforming gain of the proposed algorithm in the LOS case is 15 dB. Thus, a value of 10 dB is selected for this work as the power threshold P_{th} to judge whether the channel is a LOS channel or a NLOS channel.

4. BEAM ALIGNMENT ALGORITHM

In this section, we first formulate the beam alignment problems for both the LOS and NLOS scenarios, respectively. Next, the HBA algorithm is adopted to the considered SU-MIMO SC-FDMA system in both the LOS and NLOS scenarios to identify the optimal beam.

4.1 Problem formulation

BA is a process to identify the optimal beam code from a predefined codebook, which guarantees maximal received signal power. Since the channel behavior under the LOS and the NLOS scenarios differs significantly in the THz band, the BA problems are formulated for both cases separately in the following.

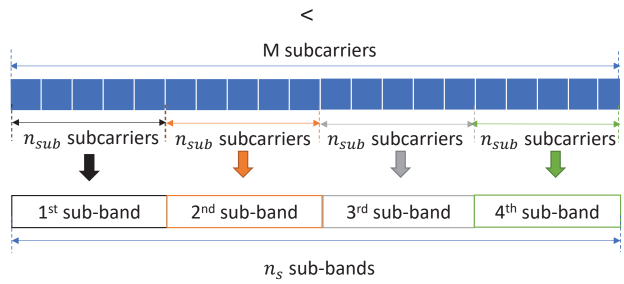


Fig. 6 – sub-band assignment for SC-FDMA system.

4.1.1 *LOS beam alignment*

In the LOS scenario, the channel behavior is dominated by the LOS propagation [23]. Hence, the beam codebook is selected as the hierarchical DFT codebook discussed in Section 3. The precoded signal at the SC-FDMA transmitter should be steered into the direction of the AoD of the LOS component ϕ_{LOS}. The corresponding spatial angle of the channel is given by $\Phi = \frac{2\pi f_c r}{c}\sin\phi_{LOS}$, where f_c is the carrier frequency of the transmission, r is the antenna spacing and c is the speed of light. For a transmission bandwidth BW, the maximal variation of the LOS spatial angle over the transmission bandwidth $\Delta\Phi$ is given by $\Delta\Phi = \frac{2\pi BW r}{c}\sin\phi_{LOS}$. For a broadband SC-FDMA system, if the same beam code is assigned to all subcarriers, some of the subcarriers may not be aligned accurately. To cope with the beam misalignment caused by a large bandwidth, all available subcarriers are divided into n_s sub-bands as shown in Fig. 6, where n_{sub} is the size of one sub-band. Here, the same beam code is assigned to all subcarriers within the same sub-band. Therefore, the maximal spatial angle variation within a sub-band $\Delta\Phi_{sub}$ should not exceed the beam coverage of the code word in the highest layer sub-codebook $\mathbf{W}^I$, i.e., $\max_{\phi_{LOS}}\Delta\Phi_{sub} \leq \frac{2\pi}{N_T}$. Hence, n_s can be determined by

$$\begin{aligned}\max_{\phi_{LOS}}\Delta\Phi_{sub} &\leq \frac{2\pi}{N_T}\\ \frac{2\pi BW\,r}{c\,n_s} &\leq \frac{2\pi}{N_T} \\ n_s &\geq \frac{BW\,r\,N_T}{c}\end{aligned} \tag{28}$$

The BA problem is equivalent to finding the optimal beam code $\mathbf{w}_i^*$ from $\mathbf{W}_{DFT}$ for the ith sub-band maximizing the mean received power over the ith sub-band. The mean received power $P_i(\mathbf{w}_i)$ for precoding vector $\mathbf{w}_i$ for ith sub-band is defined as

$$P_i(\mathbf{w}_i) = \sum_{\nu=\nu_0(i)}^{\nu_0(i)+n_{sub}-1} \mathbf{w}_i^H\mathbf{H}^H[\nu]\mathbf{H}[\nu]\mathbf{w}_i + n_{sub}N_R\sigma_n^2, \tag{29}$$

where $\nu_0(i)$ is the index of the first subcarrier in ith sub-band.

In this work, we mainly focus on the beam alignment problem for SU-MIMO transmission. Nevertheless, the proposed beam alignment framework can be adjusted to the multi-user case in a straightforward manner due to the sub-band-wise beam alignment setting. For the multi-user scenario, different users are assigned to different sub-bands such that the sub-band-wise beam alignment problem is reformulated to a multi-user beam alignment problem. This idea is similar to the frequency division multiple access in conventional communication systems. The only change here is the determination of the number of sub-bands, which should be chosen based on both the number of users and the number of transmit antennas.

4.1.2 NLOS beam alignment

In the NLOS scenario, the receive signal consists of several NLOS MPCs. To utilize the power of the NLOS com- ponents, the hierarchical k-means codebook described in Section 3 is implemented in this scenario. Due to the high frequency selectivity of the considered THz NLOS channel, even the frequency response of two neighbor- ing subcarrier may be quite distinct. Thus, the sub-band beam alignment as used for the LOS scenario is not con- sidered in the NLOS case. Instead, specific beams need to be aligned to each subcarrier used in an SC-FDMA system for a fully optimum performance. The BA problem for the νt h subcarrier is equivalent to finding the beam code $\mathbf{w}_\nu^*$ from the hierarchical k-means codebook with maximal mean receive power for the νt h subcarrier. Here, the mean receive power $P_\nu(\mathbf{w}_\nu)$ for precoding vector $\mathbf{w}_\nu$ and the νt h subcarrier is given by

$$P_\nu(\mathbf{w}_\nu) = \mathbf{w}_\nu^H \mathbf{H}^H[\nu]\mathbf{H}[\nu]\mathbf{w}_\nu + N_R\sigma_n^2. \tag{30}$$

Similar to the extension to multi-user transmission in the LOS case, the beam alignment framework for the NLOS scenario can be adjusted to a multi-user transmission as well. In the case of multi-user transmission, different users are assigned to different subcarriers such that the subcarrier-wise beam alignment algorithm is modified to a multi-user beam alignment algorithm.

4.1.3 HBA problem formulation

To accelerate the BA process in the large-scale MIMO case, we reformulate the BA problem to a stochastic MAB problem for stationary environments. The transmission system is considered as a time slotted system with T time slots to search for the optimal beam. At the begining of the BA phase, the propagation scenario can be determined at the transmitter side based on the feedback of the receive power profile. If the propagation environment is an LOS scenario, the sub-band BA discussed in the LOS beam alignment is adopted for the following BA procedure. Otherwise, the subcarrier BA introduced in the NLOS beam alignment is utilized.

If an LOS component can be received, all M subcarriers will be divided into n_s sub-bands and each sub-band has to be aligned with an optimal beam. Thus, the BA problem is transformed to n_s parallel BA problems, where n_s is the number of sub-bands. For the ith sub-band, we interpret its receive power as the measured reward r_i, i.e.,

$$r_i = \sum_{\nu=\nu_0(i)}^{\nu_0(i)+n_{sub}-1} \|\mathbf{r}[\nu]\|_2^2, \tag{31}$$

where $\mathbf{r}[\nu]$ is the receive signal vector for the νth subcarrier. Consider the beam codes in the hiearchical DFT codebook as the arms in an MAB framework. Due to the randomness of data and noise, the reward r_i is a random variable. This indicates that our LOS BA problem can be classified as a stochastic MAB problem. Assuming a stationary indoor scenario, we consider the channel as constant during the BA process. For simplicity, the reward is modeled as Gaussian distributed with variance $n_{sub}\sigma^2$. The average payoff function $\mathbb{E}[r_i(\mathbf{w}_i)]$ for a beam code $\mathbf{w}_i$ in the hierarchical DFT codebook is equivalent to the average receive power $P_i(\mathbf{w}_i)$,

$$\mathbb{E}[r_i(\mathbf{w}_i)] = P_i(\mathbf{w}_i), \tag{32}$$

In the time slot t, the transmitter selects a beam code $\mathbf{w}_i(t)$ from the hierarchical DFT codebook $\mathbf{W}_{DFT}$ for the ith sub-band. At the receiver, the power $r_i(\mathbf{w}_i(t))$ of the ith sub-band is measured and fed back to the transmitter. At the end of time slot t, the transmitter obtains the measured rewards of all sub-bands for this time slot and decides which arm to select for which sub-band based on a specific rule for the next time slot $t+1$.

For a stochastic MAB problem, the performance of the algorithm is evaluated via the expected cumulative regrets over T time slots. Here, the expected cumulative regrets for the ith sub-band $R_i(T)$ is defined as the expected difference between the cumulative reward of the optimal arm $\mathbf{w}_i^*$ and the cumulative reward of the proposed algorithm for ith sub-band, given by

$$R_i(T) = \sum_{t=1}^{T} \left(r_i\left(\mathbf{w}_i^*\right) - r_i\left(\mathbf{w}_i(t)\right)\right). \tag{33}$$

The objective of the design of the MAB algorithm is to find a selection policy that minimizes the sum expected cumulative regret $R_{LOS}(T)$ over all sub-bands, i.e., $R_{LOS}(T) = \sum_{i=1}^{n_s} R_i(T)$.

The BA problem for the NLOS scenario is similar to that for the LOS scenario. Here, the BA problem can be transformed to M parallel BA problem. For the νth subcarrier, we interpret its receive power $\|\mathbf{r}[\nu]\|_2^2$ as the measured reward r_ν, and consider the beam codes in the hierarchical k-means codebook as the arms in an MAB framework. The reward is modeled as Gaussian distributed with variance σ^2. The average payoff function $\mathbb{E}[r_\nu(\mathbf{w}_\nu)]$ for beam code $\mathbf{w}_\nu$ is equivalent to the average receive power $P_\nu(\mathbf{w}_\nu)$,

$$\mathbb{E}[r_\nu(\mathbf{w}_\nu)] = P_\nu(\mathbf{w}_\nu) = \sigma_a^2 \mathbf{w}_\nu^H \mathbf{H}^H[\nu]\,\mathbf{H}[\nu]\,\mathbf{w}_\nu + N_r\sigma_n^2. \tag{34}$$

In each time slot t, the transmitter selects beam codes for each subcarrier from the hierarchical k-means codebook. At the receiver, the receive power r_ν for the νth subcarrier is measured and fed back to the transmitter. At the end of time slot t, the transmitter obtains the measured rewards of all subcarriers for this time slot and decides which arm to select for which subcarrier based on a specific rule for the next time slot $t+1$. In the NLOS scenario, the expected cumulative regrets $R_\nu(T)$ for νth subcarrier is given by

$$R_\nu(T) = \sum_{t=1}^{T} \left(r_\nu\left(\mathbf{w}_\nu^*\right) - r_\nu\left(\mathbf{w}_\nu(t)\right)\right). \tag{35}$$

where $\mathbf{w}_\nu^*$ is the optimal beam code for the νth subcarrier. The objective of the algorithm design for the NLOS case is to find a selection policy that minimizes the sum expected cumulative regret $R_{NLOS}(T)$ over all subcarriers, i.e., $R_{NLOS}(T) = \sum_{\nu=\nu_0}^{\nu_0+M-1} R_\nu(T)$.

4.2 HBA procedure

In the following, we extend the HBA algorithm proposed in [9] to the considered LOS and NLOS SC-FDMA transmission. As mentioned, the BA problems for both LOS and NLOS scenarios, respectively, can be decomposed into several independent sub-BA problems, which can be solved in parallel. The HBA algorithm discussed in the following provides an efficient approach to solve one independent BA problem, such as a sub-band BA problem in the LOS case and a subcarrier BA problem in the NLOS case, respectively.
The procedure of the HBA algorithm is shown in Algorithm 2. The algorithm is designed based on the hierarchical structure of the average payoff function with respect to the code words. For a single BA problem, consider the beam code $\mathbf{w}(i,j)$ in the hierarchical codebook. If $\mathbf{w}(i,j)$ performs well, its left child $\mathbf{w}(i+1, 2j-1)$ and right child $\mathbf{w}(i+1, 2j)$ are highly likely to perform well, too. Hence, the algorithm will explore intensively within the beam coverage $CV(\mathbf{w}(i,j))$ with a good reward and loosely in others. For this proposal, a search tree is generated with nodes associated with the beam coverage, independently for each sub-BA problem. The node in a higher layer covers a smaller region as discussed in the codebook design part. The algorithm operates in discrete time slots, and in each time slot t, a new beam code is chosen by a deterministic rule based on the attributes of the search tree. The code selection procedure is executed independently for each sub-band or subcarrier in the LOS or NLOS scenario, respectively. After measurement at the receiver, the observed rewards will be fed back to the transmitter. Afterwards, the transmitter updates the attributes of the entire search tree based on the newly observed rewards r^t and the newly selected code is added to the search tree. By repeating this selection-update process, the HBA algorithm narrows the searching region intelligently until a close-to-optimal beam code is selected.

Algorithm 2 HAB Procedure in SC-FDMA

Input: $\rho_1, \gamma, \sigma^2, t_{max}$
Output: $\mathbf{b}^*$

1: Initialization: Set $\mathbf{T}^t = \{(1,1)\}$, $Q_{2,1} = Q_{2,2} = +\infty$ and $b^t = (1,1)$;
2: **while** $b^t \neq b^{t-1}$ && $t < t_{max}$ **do**
3: $\quad h = 1, j = 1$ and $\mathbb{P} = \{(h,j)\}$;
4: $\quad$ *New Node Selection*
5: $\quad$ **while** $(h,j) \in \mathbf{T}^t$ **do**
6: $\qquad$ Select new node (h,j) based on *(36)*;
7: $\qquad$ $\mathbb{P} = \mathbb{P} \cup (h,j)$;
8: $\quad$ **end while**
9: $\quad$ Selected new code: $(H^t, J^t) = (h,j)$;
10: $\quad$ Decision Tree Update: $\mathbf{T}^{t+1} = \mathbf{T}^t \cup (H^t, J^t)$;
11: $\quad$ Measure the reward r^t;
12: $\quad$ *Attribute Update*
13: $\quad$ **for all** $(h,j) \in \mathbf{T}^t$ **do**
14: $\qquad$ **if** $(h,j) \in \mathbb{P}$ **then**
15: $\qquad\quad$ update $N_{h,j}(t)$, $R_{h,j}(t)$ and $E_{h,j}(t)$ with *(37)*, *(38)* and *(39)*, respectively;
16: $\qquad$ **end if**
17: $\qquad$ update $E_{h,j}(t)$ with *(39)*;
18: $\quad$ **end for**
19: $\quad$ $Q_{H^t+1,2J^t-1}(t) = Q_{H^t+1,2J^t-1}(t) = +\infty$;
20: $\quad$ **for all** $(h,j) \in \mathbf{T}^t$ **do**
21: $\qquad$ update $Q_{h,j}(t)$ with *(40)*;
22: $\quad$ **end for**
23: **end while**

New node selection
In the new node selection phase, the HBA will select the beam code $\mathbf{w}^t$ with maximal estimated mean reward—Q-value for one subcarrier or sub-band. The procedure of node selection can be found from line 3 to line 10 in Algorithm 2. By exploiting the hierarchical codebook structure, a binary tree search is implemented to find the new node. The binary search tree $\mathbf{T}$ is initialized in the beginning. A node in $\mathbf{T}$ is represented by (h,j), where h is the depth from the root node and $j, 1 \leq j \leq 2^{h-1}$ is the index at depth h. The corresponding beam code of node (h,j) is $\mathbf{w}(h,j)$. After initialization, $\mathbf{T}^1$ contains only the root $(1,1)$. Assume $\mathbf{T}^t$ is the search tree at time t. Starting from the root node, the Q-values of two child nodes are compared until a new node $(H^t, J^t) \notin \mathbf{T}^t$ is selected. For a node (h,j), the selection criterion is to choose its child $(h+1, j^*)$ with largest Q-value, that is:

$$j^* = \begin{cases} 2j, & Q_{h+1,2j}(t) > Q_{h+1,2j-1}(t) \\ 2j-1, & Q_{h+1,2j}(t) < Q_{h+1,2j-1}(t) \\ 2j - \text{Ber}(0.5), & Q_{h+1,2j}(t) = Q_{h+1,2j-1}(t) \end{cases} \tag{36}$$

where $\text{Ber}(0.5)$ represents a Bernoulli distributed random variable with a parameter of 0.5. All selected nodes during the compare-select procedure are saved in $\mathcal{P}$, which is the path from the root node to the selected node. The finally selected node (H^t, J^t) is added to the search tree $\mathbf{T}^t$ to obtain the search tree $\mathbf{T}^{t+1}$ for the next time slot $t+1$, $\mathbf{T}^{t+1} = \mathbf{T}^t \cup (H^t, J^t)$.

Attribute update

In this step, the attributes of all nodes in the search tree $\mathbf{T}^{t+1}$ are updated based on the feedback of the measured reward r^t in the current time slot. For LOS BA, r^t is the measured receive power of one sub-band, while r^t is the measured receive power of one subcarrier in the NLOS case. The details of the attribute update are shown in Algorithm 2 from line 12 to line 22. The update of Q-values consists of the following steps.

At first, the number of times $N_{h,j}(t)$ that node (h, j) has been visited until time slot t is updated as

$$N_{h,j}(t) = N_{h,j}(t-1) + 1, \forall (h,j) \in \mathcal{P}_\nu. \quad (37)$$

Node (h, j) must have been visited one time when it is selected as the new node for the search tree. (h, j) will be visited one more time when one of its descendants is added to the search tree $\mathbf{T}^t$. The average measured reward $R_{h,j}$ of (h, j) is updated by

$$R_{h,j}(t) = \frac{\left(N_{h,j}(t) - 1\right) R_{h,j}(t-1) + r^t}{N_{h,j}(t)}, \forall (h,j) \in \mathcal{P}. \quad (38)$$

The empirical average reward $E_{h,j}(t)$ of node (h, j) in time slot t is defined as

$$E_{h,j}(t) = \begin{cases} R_{h,j}(t) + \sqrt{\frac{2\sigma^2 \log t}{N_{h,j}(t)}} + \rho_1 \gamma^h, & \text{if } N_{h,j}(t) > 0 \\ +\infty, & \text{otherwise} \end{cases} \quad (39)$$

where $\sqrt{\frac{2\sigma^2 \log t}{N_{h,j}(t)}}$ represents the confidence margin related to the uncertainty of rewards, related to random data and noise. With increasing $N_{h,j}(t)$, the uncertainty of the reward of (h, j) becomes lower, since there are more available observations. The confidence margin is designed based on Bayesian principle and derived in [26]. Here, $0 < \gamma < 1$ and $\rho_1 > 0$ are parameters of the algorithm, and $\rho_1\gamma^h$ specifies the maximum variation of the average reward function within beam coverage CV $(\mathbf{w}(h, j))$ [26], which depend on the codebook structure. The datails re- garding γ and ρ selection can be found in [26] and [9]. If γ and ρ is chosen based on the bounded diameter prin- ciple and well-shaped region principle from [9], HBA will converge to the optimal beam code with high probability. In the initial phase of the HBA, no information regarding the rewards is available. Hence, $E_{h,j}(t)$ is initialized by infinity. With abundant observed rewards within $CV\,(h, j)$ available, we can tighten the upperbound of mean rewards step by step. Finally, the estimated maximum mean reward $Q(h, j)$ in beam coverage $CV(h, j)$ is determined as [26]

$$Q_{h,j}(t) = \begin{cases} \min\{E_{h,j}(t), \\ \max\left\{Q_{h+1,2j-1}(t), Q_{h+1,2j}(t)\right\}\}, \\ \text{ if } N_{h,j}(t) > 0 \\ +\infty, \text{ otherwise} \end{cases} \quad (40)$$

When the HBA algorithm has obtained a sufficient number of observed rewards within the searching tree, it will narrow its searching coverage in the highest layer. The algorithm will be terminated, if no new node has been selected and the selection result no longer changes, i.e.,

$$\mathbf{T}^{t+1} = \mathbf{T}^t. \quad (41)$$

Then, the currently selected beam $\mathbf{w}(H^t, J^t)$ is the derived beam for the correponding sub-band and subcarrier in LOS and NLOS cases, respectively. According to [9], the computational complexity of the HBA is quadratic in the number of processed time slots, $O(T^2)$.

Table 1 – SC-FDMA system setting

Parameter	Numerical Value
DFT size M	1200
DFT size N	2048
Cyclic prefix length L_c	512
Transmission band	$0.1 - 0.1281$ THz
Number of transmit antennas N_t	64
Number of receive antennas N_r	2
Antenna gain $G_t, G_R, G_t = G_r$	20 dBi
Signal constellation	4QAM

4.3 Complexity analysis

At time slot T, for one subcarrier or sub-band, the decision tree contains T nodes as the tree is extended by one node in each time slot. The attributes of all nodes in a decision tree should be updated in each time slot, and hence the run time in time slot T is linear in T, i.e., $O(T)$. As the algorithm is executed for T time slots, the total computational complexity of the proposed HBA algorithm is quadratic in T , i.e., $O(T^2)$ [9].

5. NUMERICAL RESULTS

In the following, we investigate a THz SC-FDMA system, whose parameter settings are provided in Table 1. The transmission scenario is the indoor scenario considered in [6]. The transmitter is fixed at the center of the room ceiling and the location of the receiver with fixed height $h = 1.5$ m is uniformly distributed within the indoor environment. The results are averaged over 500 channel realizations. The proposed algorithm is compared to the following benchmarks:

Optimal SC-FDMA beamforming: In this beamforming scheme, the CSI is considered as known at both the receiver and the transmitter. Thus, an MMSE frequency domain equalizer according to [10] can be designed. This algorithm aims to minimize the MSE after equalization, which can be formulated as a convex optimization problem. The optimal solution is derived in [10], and its performance can be regarded as a performance upper bound for our proposed scheme.

Random beamforming: In random beamforming, the beamforming vector for each subcarrier or sub-band is a random complex vector with constant L_2-norm and random phase profile. An MMSE equalizer is employed at the

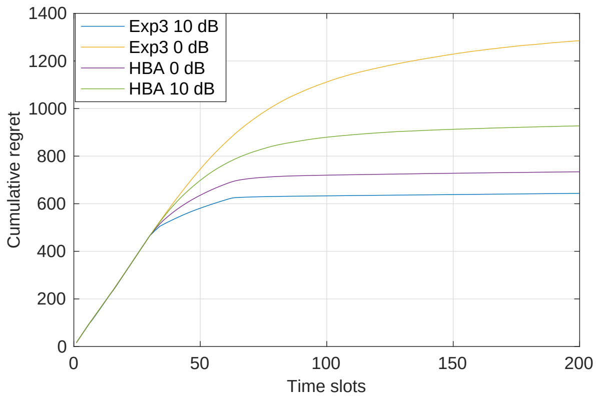


Fig. 7 – Cumulative regret of different algorithms in LOS scenario.

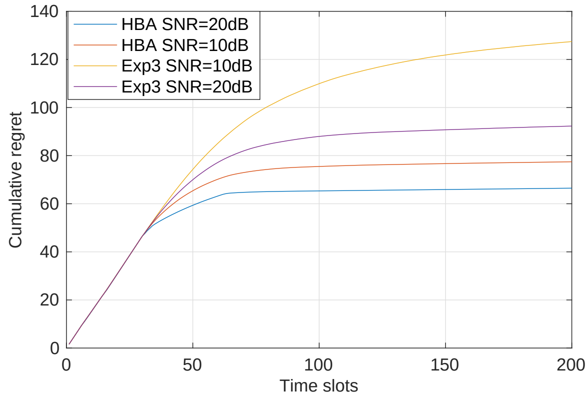


Fig. 8 – Cumulative regret of different algorithms in NLOS scenario.

receiver as well. This scheme is related to a performance lower bound.

Exhaustive search: Exhaustive search is a naive beam alignment approach. In this scheme, the transmitter applies all beam codes from the predefined codebook several times to obtain the rewards of all beam codes [9]. Then the beam code with the largest measured reward is selected for usage. This beamforming method ensures that the optimal beam code from the available codebook is always obtained. Its performance serves for quantifying the loss due to the beam misalignment caused by the HBA.

Exponential weights (Exp3) algorithm: The Exp3 algorithm is based on the adversarial MAB framework. In [16], the authors advocated applying the Exp3 algorithm to the beam alignment problem. Compared to HBA, the Exp3 algorithm does not take the hierarchical structure of the codebook into account, which results in a slow convergence behavior for the beam code selection. Its convergence behavior can be taken as a reference for the convergence behavior of the HBA.

5.1 Cumulative regret

Figures 7 and 8 depict the sum cumulative regret $R(T)$ performance of the proposed HBA algorithm in LOS and NLOS scenarios, respectively, where the curves have been averaged over the receiver locations. Here, SNR is defined as the transmit power of one antenna divided by the noise power at one receive antenna. First, a bounded regret behavior is observed for both LOS and NLOS scenarios, which complies with the conclusion from [9]. In addition, the cumulative regret and the noise power are positively correlated, which indicates that the HBA needs more time slots to converge to the optimal beam under low SNR. However, under all SNR conditions, the HBA can achieve nearly 100% beam accuracy after 40 time slots, as confirmed by the bounded regret behavior. Moreover, under all SNR conditions in both LOS and NLOS scenarios, HBA performs better than Exp3 with respect to cumulative regrets. This is due to the fact that HBA utilizes the hierarchical structure of the codebook. Thus, searching over the entire codebook is avoided in HBA. Meanwhile, Exp3 operates like a random searching in the beginning, which results in a large number of time slots to converge.

5.2 Convergence behavior

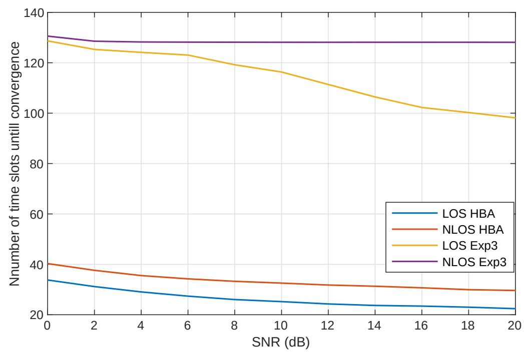


Fig. 9 – Convergence behavior of different BA schemes in both LOS and NLOS scenario.

Figure 9 illustrates the convergence behavior of the HBA in both LOS and NLOS scenario. In high SNR conditions, only 30 time slots are required for convergence to the finally selected beam, which is much faster compared to the Exp3 algorithm, requiring nearly 100 time slots to converge [9]. Furthermore, the convergence speed is related to the SNR, i.e., the HBA needs more time slots to converge to the optimal beam under low SNR, suggested also by Fig. 9. The reason is that in low SNR, the measured rewards are severely affected by the noise and the HBA requires more feedback information from the receiver to determine the mean reward. In all SNR conditions, our proposed approach converges faster compared to the benchmark schemes.

5.3 BER performance

In Fig. 10, the BER performance of the different beamforming schemes is shown for the LOS scenario. Apparently, the performance of the HBA is significantly better than that of random beamforming. Furthermore, the HBA

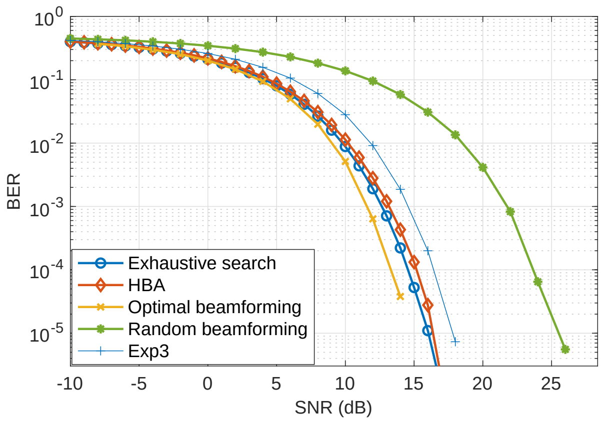


Fig. 10 – Average BER vs. SNR for different beamforming schemes in LOS scenario.

achieves a performance close to that of the optimal beamforming and exhaustive search, respectively, which complies with Theorem 2 and demonstrates the benefits of the HBA. In addition, although the Exp3 algorithm can achieve a similar BER performance as the HBA, Exp3 requires double the number of the time slots than HBA. Hence, HBA is able to provide a close-to-optimal beam selection with significantly shorter latency in the LOS scenario compared to the benchmark schemes.

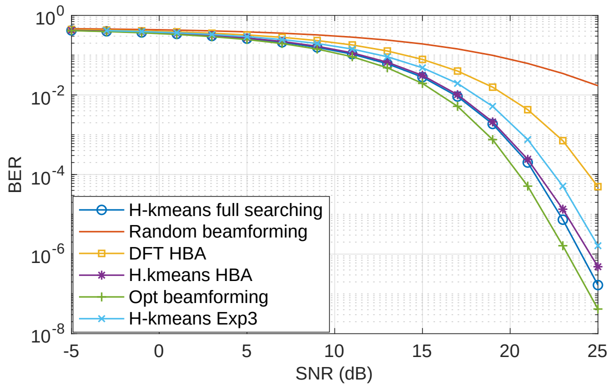


Fig. 11 – Average BER vs. SNR for different beamforming schemes in NLOS scenario.

In Fig. 11, the BER performance for the NLOS scenario is shown. To illustrate the performance of a hierarchical k-means codebook, the performance of the HBA with hierarchical DFT codebook is shown in addition as a benchmark. First, HBA with hierarchical k-means codebook design can improve the system performance by 3 dB compared to the HBA with DFT codebook. The HBA algorithm in the NLOS scenario achieves a performance close to the performance upper bound as well. The gap between the proposed scheme and optimal beamforming is reduced to less than 1 dB. Thus, the hierarchical k-means codebook design can better exploit the NLOS components compared to the DFT codebook. We can also state that HBA with hierarchical k-means codebook is more likely to converge to a suboptimal beam code than HBA in the LOS case, according to the gap between exhaustive search and HBA. The reason is that the NLOS channel suffers from strong frequency-selectivity. Thus, the training channels within the same beam coverage may have a large distance, which might direct the HBA to a non-optimal code. Besides, the reduced channel quality in the NLOS case increases the misalignment rate significantly. However, this performance degradation is less than 1 dB, which is acceptable compared to the high exploration cost for the exhaustive search.

6. CONCLUSION

In this paper, hierarchical beam alignment with hierarchical codebook design for SU-MIMO THz communications has been studied. First, the hierarchical codebook design problem in MIMO THz communications has been established. Next, the hierarchical codebooks for LOS and NLOS propagation have been designed based on DFT codebook and data-driven hierarchical k-means clustering, respectively. Then, the beam alignment problem in THz communications has been formulated and the HBA from mmWave communications is adjusted to the SC-FDMA SU-MIMO THz communication system. Numerical results show that HBA combined with hierarchical DFT codebook can achieve a performance close to the optimal beamforming from [10] in a LOS scenario, while in an NLOS scenario HBA combined with hierarchical k-means codebook outperforms the DFT codebook. In our future work, the HBA will be extended to a multi-user transmis- sion.

APPENDIX A

Proof of Theorem 1

Here, the maximum over the codebook can be replaced by the L_∞-norm as

$$\int_\phi \max_{\mathbf{w}\in\mathbf{W}^I} \|\mathbf{a}^T(N_T,\phi)\mathbf{w}\|_2^2\, d\phi = \int_\phi \|\mathbf{a}^T(N_T,\phi)\mathbf{W}^I\|_\infty^2 d\phi \tag{42}$$

Regarding the optimization problem in (14), the constraints can be relaxed to a convex constraint, i.e., $\mathbf{w}_j^H\mathbf{w}_j \le 1, 1\le j\le 2^{i-1}$, resulting in

$$\begin{aligned} &\max_{\mathbf{W}} \int \|\mathbf{a}^T(N_T,\phi)\mathbf{W}\|_\infty^2 \mathrm{d}\phi \\ &\text{s.t. } \mathbf{w}_j^H\mathbf{w}_j \le 1, i=1,2,\cdots,N_T, 1\le j\le N_T. \end{aligned} \tag{43}$$

$\mathbf{F}$ is a local optimum for (12), if and only if there exists a Lagrange multiplier vector λ guarantees the KKT conditions are satisfied. The Lagrangian of (12) is given by

$$\hat{G}(\mathbf{W},\lambda) = \int_\phi \|\mathbf{a}^T(N_t,\phi)\mathbf{W}\|_\infty^2 d\phi - \sum_{i=1}^{N_T}\lambda_j(\mathbf{w}_j^H\mathbf{w}_j - 1). \tag{44}$$

The gradient of Lagrangian with respect to $F_{i,k}^*$ is given by

$$\begin{aligned}\frac{d\hat{G}}{dF_{i,k}^*} =& \frac{d\int_0^{2\pi}(\|\mathbf{Fa}(N_t,\phi)\|_\infty^2)d\phi}{dF_{i,k}^*} - \lambda_k F_{i,k} \\ =& \int_{\frac{2i-1}{N_T}\pi}^{\frac{2i+1}{N_T}\pi} \frac{d|\sum_{l=0}^{N_T-1} F_{i,l} e^{j(l-1)\phi}|^2}{dF_{i,k}^*} d\phi - \lambda_k F_{i,k} \\ =& \int_{\frac{2i-1}{N_T}\pi}^{\frac{2i+1}{N_T}\pi} \sum_{l=0}^{N_T-1} e^{\frac{-j2\pi il}{N_T}} e^{j\phi(l-k)} d\phi - \lambda_k F_{i,k} \\ =& e^{\frac{-j2\pi ik}{N_T}} \sum_{l=0}^{N_T-1} e^{\frac{-j2\pi i(l-k)}{N_T}} \frac{e^{\frac{(l-k)(2i+1)}{N_T}} - e^{\frac{(l-k)(2i-1)}{N_T}}}{j(i-k)} \\ & - \lambda_k F_{i,k} \\ =& e^{\frac{-j2\pi ik}{N_T}} \sum_{l=0}^{N_T-1} \frac{2\sin((l-k)\pi)}{l-k} - \lambda_k F_{i,k} \\ =& \hat{D}_{i,k} F_{i,k} - \lambda_k F_{i,k},\end{aligned} \tag{45}$$

where $\hat{D}_{i,k} = \sum_{l=0}^{N_T-1} \frac{2\sin((l-k)\pi)}{l-k}$, which is not dependent on i. Therefore, there exists a Lagrange multiplier λ that satisfies the KKT conditions. Hence, $\mathbf{F}$ is a local optimum for the optimization problem (14) and (42).

APPENDIX B

Proof of Theorem 2

Proof. According to [16], the receive power $|\mathbf{a}_s^H(N_t,\Phi)\mathbf{a}_s(N_t,\Phi_{i,j})|^2$ is given by

$$|\mathbf{a}_s^H(N_t,\Phi)\mathbf{a}_s(N_t,\Phi_{i,j})|^2 = \frac{\sin^2(N_t(\Phi_{i,j}-\Phi))}{\sin^2(\Phi_{i,j}-\Phi)} \tag{46}$$

Inserting $\Phi_{i,j} = \frac{\pi(2j-1)2^{I-i}}{N_t}$ to the above equation, we obtain:

$$\begin{aligned}\frac{\sin^2(N_t(\Phi_{i,j}-\Phi))}{\sin^2((\Phi_{i,j}-\Phi))} &= \frac{\sin^2(N_t(\frac{\pi(2j-1)2^{I-i}}{N_t}-\Phi))}{\sin^2(\Phi_{i,j}-\Phi)} \\ &= \frac{\sin^2(N_t\Phi)}{\sin^2(\Phi_{i,j}-\Phi)}.\end{aligned} \tag{47}$$

Hence, maximizing the receive power $|\mathbf{a}_s^H(N_t,\Phi)\mathbf{a}_s(N_t,\Phi_{i,j})|^2$ over the codebook in the i^{th} layer is equivalent to finding the minimum of $\sin^2(\Phi_{i,j}-\Phi)$, which corresponds to minimizing $|\Phi_{i,j}-\Phi|$ with $1 \le j \le 2^{i-1}$. In other words, the beam coverage of $\mathbf{w}(i,j)$ is a set of the steering vector with LOS spatial angle close to $\Phi_{i,j}$. Since the distance between $\Phi_{i,j}$ and $\Phi_{i,j-1}$ is $\frac{4\pi}{2^i}$, the beam coverage of $\mathbf{w}(i,j)$ is obtained as

$$CV(\mathbf{w}(i,j)) = \left[\Phi_{i,j} - \frac{2\pi}{2^i}, \Phi_{i,j} + \frac{2\pi}{2^i}\right]. \tag{48}$$

Next, N_t is set to 2^I. By inserting $N_t = 2^I$ and $\Phi_{i,j} = \frac{\pi(2j-1)2^{I-i}}{N_t}$ into the above equation, the result in (17) can be obtained. □

APPENDIX C

Proof of Theorem 3

The objective function in *(22)* can be reformulated as

$$\begin{aligned}&\sum_{i=1}^{N_p}\sum_{\nu=\nu_0}^{\nu_0+M-1} \max_{\mathbf{w}_j\in\mathbf{W}} \|\mathbf{H}_{NLOS}^{\nu,i}\mathbf{w}_j\|_2^2 \\ =&\sum_{i=1}^{N_p}\sum_{\nu=\nu_0}^{\nu_0+M-1} \max_{\mathbf{w}_j\in\mathbf{W}} \mathrm{tr}((\mathbf{H}_{NLOS}^{\nu,i})^H\mathbf{H}_{NLOS}^{\nu,i}\mathbf{w}_j\mathbf{w}_j^H) \\ =&\sum_{i=1}^{N_p}\sum_{\nu=\nu_0}^{\nu_0+M-1} \max_{\mathbf{w}_j\in\mathbf{W}} \\ &-\frac{1}{2}(\mathrm{tr}((\hat{\mathbf{H}}_{\nu,i}-\hat{\mathbf{W}}_j)(\hat{\mathbf{H}}_{\nu,i}-\hat{\mathbf{W}}_j)^H)) \\ &+\frac{1}{2}\mathrm{tr}(\hat{\mathbf{H}}_{\nu,i}\hat{\mathbf{H}}_{\nu,i}^H) + \frac{1}{2}\mathrm{tr}(\hat{\mathbf{W}}_j\hat{\mathbf{W}}_j^H),\end{aligned} \tag{49}$$

where $\hat{\mathbf{H}}_{\nu,i}$ and $\hat{\mathbf{W}}_j$ are defined as $(\mathbf{H}_{NLOS}^{\nu,i})^H\mathbf{H}_{NLOS}^{\nu,i}$ and $\mathbf{w}_j\mathbf{w}_j^H$, respectively. Since $\mathrm{tr}(\hat{\mathbf{H}}_{\nu,i}\hat{\mathbf{H}}_{\nu,i}^H)$ and $\mathrm{tr}(\hat{\mathbf{W}}_j\hat{\mathbf{W}}_j^H)$ are constant, the optimization problem in *(22)* is equivalent to

$$\min_{\mathbf{W}} \sum_{i=1}^{N_p}\sum_{\nu=\nu_0}^{\nu_0+M-1} \min_{\mathbf{w}_j\in\mathbf{W}} D(\hat{\mathbf{H}}_{\nu,i},\hat{\mathbf{W}}_j) \tag{50}$$

APPENDIX D

Proof of Theorem 4

Since $rank(\mathbf{W})$ is fixed to 1, $\mathbf{W}$ can always be decomposed as

$$\mathbf{W} = \mathbf{w}_x\mathbf{w}_x^H. \tag{51}$$

Hence, the optimization problem in *(26)* is equivalent to

$$\begin{aligned}\max_{\mathbf{w}_x} \quad & \sum_{\mathbf{H}\in\mathcal{H}_j} \mathbf{w}_x^H\mathbf{H}\mathbf{w}_x \\ \text{s.t.} \quad & \mathbf{w}_x^H\mathbf{w}_x = 1.\end{aligned} \tag{52}$$

The optimal solution for $\mathbf{w}_x$ is given by $\mathbf{w}_{max}$, where $\mathbf{w}_{max}$ is the eigenvector of $\sum_{\mathbf{H}\in\mathcal{H}_j}\mathbf{H}$ corresponding to its largest eigenvalue.

REFERENCES

[1] Ian F Akyildiz, Josep Miquel Jornet, and Chong Han. "Terahertz band: Next frontier for wireless communications". In: *Physical Communication* 12 (2014), pp. 16–32.

[2] Shahid Mumtaz, Josep Miquel Jornet, Jocelyn Aulin, Wolfgang H Gerstacker, Xiaodai Dong, and Bo Ai. "Terahertz communication for vehicular networks". In: *IEEE Transactions on Vehicular Technology* 66.7 (2017).

[3] Thomas Kürner. "Towards future THz communications systems". In: *Terahertz science and technology* 5.1 (2012), pp. 11–17.

[4] Zhi Chen, Xinying Ma, Bo Zhang, Yaxin Zhang, Zhongqian Niu, Ningyuan Kuang, Wenjie Chen, Lingxiang Li, and Shaoqian Li. "A survey on terahertz communications". In: *China Communications* 16.2 (2019), pp. 1–35.

[5] Viktoria Schram, Yifei Wu, Monika Kolleshi, and Wolfgang Gerstacker. "Comparison of Transmission Concepts for Indoor THz Communication Systems". In: *2020 Third International Workshop on Mobile Terahertz Systems (IWMTS)*. IEEE. 2020, pp. 1–5.

[6] Anamaria Moldovan, Michael A Ruder, Ian F Akyildiz, and Wolfgang H Gerstacker. "LOS and NLOS channel modeling for terahertz wireless communication with scattered rays". In: *2014 IEEE Globecom Workshops (GC Wkshps)*. IEEE. 2014, pp. 388–392.

[7] Hyung G Myung and David J Goodman. *Single carrier FDMA: a new air interface for long term evolution*. Vol. 8. John Wiley & Sons, 2008.

[8] Chong Han, A Ozan Bicen, and Ian F Akyildiz. "Multi-ray channel modeling and wideband characterization for wireless communications in the terahertz band". In: *IEEE Transactions on Wireless Communications* 14.5 (2014), pp. 2402–2412.

[9] Wen Wu, Nan Cheng, Ning Zhang, Peng Yang, Weihua Zhuang, and Xuemin Shen. "Fast mmWave beam alignment via correlated bandit learning". In: *IEEE Transactions on Wireless Communications* 18.12 (2019), pp. 5894–5908.

[10] Uyen Ly Dang, Michael A Ruder, Robert Schober, and Wolfgang H Gerstacker. "MMSE beamforming for SC-FDMA transmission over MIMO ISI channels". In: *EURASIP Journal on Advances in Signal Processing* 2011 (2011), pp. 1–11.

[11] Viktoria Schram, Anamaria Moldovan, and Wolfgang H Gerstacker. "Compressive sensing for indoor THz channel estimation". In: *2018 52nd Asilomar Conference on Signals, Systems, and Computers*. IEEE. 2018, pp. 1539–1546.

[12] Chunxiao Jiang, Haijun Zhang, Yong Ren, Zhu Han, Kwang-Cheng Chen, and Lajos Hanzo. "Machine learning paradigms for next-generation wireless networks". In: *IEEE Wireless Communications* 24.2 (2016), pp. 98–105.

[13] Nikhil Gulati and Kapil R Dandekar. "Learning state selection for reconfigurable antennas: A multi-armed bandit approach". In: *IEEE Transactions on Antennas and Propagation* 62.3 (2013), pp. 1027–1038.

[14] Morteza Hashemi, Ashutosh Sabharwal, C Emre Koksal, and Ness B Shroff. "Efficient beam alignment in millimeter wave systems using contextual bandits". In: *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*. IEEE. 2018, pp. 2393–2401.

[15] Jianjun Zhang, Yongming Huang, Yu Zhou, and Xiaohu You. "Beam Alignment and Tracking for Millimeter Wave Communications via Bandit Learning". In: *IEEE Transactions on Communications* (2020).

[16] Irched Chafaa, E Veronica Belmega, and Mérouane Debbah. "Adversarial multi-armed bandit for mmwave beam alignment with one-bit feedback". In: *Proceedings of the 12th EAI International Conference on Performance Evaluation Methodologies and Tools*. 2019, pp. 23–30.

[17] Jianhua Mo, Boon Loong Ng, Sanghyun Chang, Pengda Huang, Mandar N Kulkarni, Ahmad AlAmmouri, Jianzhong Charlie Zhang, Jeongheum Lee, and Won-Joon Choi. "Beam codebook design for 5G mmWave terminals". In: *IEEE Access* 7 (2019), pp. 98387–98404.

[18] Ken Kreutz-Delgado. "The complex gradient operator and the CR-calculus". In: *arXiv preprint arXiv:0906.4835* (2009).

[19] Radoslaw Piesiewicz, Christian Jansen, Daniel Mittleman, Thomas Kleine-Ostmann, Martin Koch, and Thomas Kurner. "Scattering analysis for the modeling of THz communication systems". In: *IEEE Trans- actions on Antennas and Propagation* 55.11 (2007), pp. 3002–3009.

[20] Sebastian Priebe and Thomas Kurner. "Stochastic modeling of THz indoor radio channels". In: *IEEE Transactions on Wireless Communications* 12.9 (2013), pp. 4445–4455.

[21] Zhengqing Yun, Zhijun Zhang, and Magdy F Iskander. "A ray-tracing method based on the triangular grid approach and application to propagation prediction in urban environments". In: *IEEE Transactions on Antennas and Propagation* 50.5 (2002), pp. 750–758.

[22] Y. Chen, Y. Li, C. Han, Z. Yu, and G. Wang. "Channel Measurement and Ray-Tracing-Statistical Hybrid Modeling for Low-Terahertz Indoor Communications". In: *arXiv e-prints* (2021).

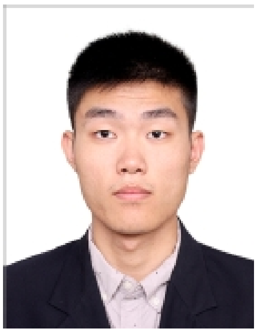
[23] Cen Lin and Geoffrey Ye Li. "Indoor terahertz communications: How many antenna arrays are needed?" In: IEEE Transactions on Wireless Communications 14.6 (2015), pp. 3097–3107.

[24] Laura Pometcu and Raffaele D'Errico. "Channel Model Characteristics in D-Band for NLOS Indoor Scenarios". In: *2019 13th European Conference on Antennas and Propagation (EuCAP)*. IEEE. 2019, pp. 1–4.

[25] Zhenyu Xiao, Tong He, Pengfei Xia, and Xiang-Gen Xia. "Hierarchical codebook design for beamforming training in millimeter-wave communication". In: *IEEE Transactions on Wireless Communications* 15.5 (2016), pp. 3380–3392.

[26] Sébastien Bubeck, Gilles Stoltz, Csaba Szepesvári, and Rémi Munos. "Online optimization in X-armed bandits". In: *Advances in Neural Information Processing Systems.* 2009, pp. 201–208.

AUTHORS



Yifei Wu received the B.Eng. degree in automation from Zhejiang University, Hangzhou, China, in 2017, and the M.S. degree in electrical engineering from Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, in 2020, where he is currently pursuing the Ph.D. degree. His research interests include terahertz communication, intelligent reflecting surface-assisted communications, and convex optimization.

Johannes Koch was born in Bochum, Germany, in 1992. He received the M.Sc. degree in electronic engineering from the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany in 2018, where he is currently pursuing the PH.D. degree. In 2018 he joined the Institute for Microwaves and Photonics (LHFT), FAU. His current research interests include THz communication systems, THz channel measurements and channel modeling above 100 GHz.

Martin Vossiek received the Ph.D. degree from Ruhr-Universität Bochum, Bochum, Germany, in 1996. In 1996, he joined Siemens Corporate Technology, Munich, Germany, where he was the Head of the Microwave Systems Group, from 2000 to 2003. Since 2003, he has been a Full Professor with Clausthal University, Clausthal-Zellerfeld, Germany. Since 2011, he has been the Chair of the Institute of Microwaves and Photonics (LHFT), Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany. He has authored or coau- thored more than 300 articles. His research has led to over 90 granted patents. His current research in- terests include radar, transponder, RF identification, communication, and wireless locating systems. Martin Vossiek is a member of the German National Academy of Science and Engineering (acatech) and of the German research foundation (DFG) review board. He is a member of the German IEEE Microwave Theory and Techniques (MTT)/ Antennas and Propagation (AP) Chapter Executive Board and a member of the IEEE MTT Technical Committees MTT-24 Microwave/mm-wave Radar, Sensing, and Array Systems; MTT-27 Connected and Autonomous Systems (as founding chair); MTT-29 Microwave Aerospace Systems. He is also serving on the advisory board of the IEEE CRFID Technical Committee on Motion Capture & Localization. Martin Vossiek has received more than 10 best paper prices and several other awards. For example, he was awarded the 2019 Microwave Application Award from the IEEE MTT Soci- ety (MTT-S) for Pioneering Research in Wireless Local Positioning Systems. Dr. Vossiek has been a member of organizing committees and technical program com- mittees for many international conferences and he has served on the Review Boards for numerous technical journals. From 2013 to 2019, he was an Associate Editor for the IEEE TRANSACTIONS ON MTT.

Wolfgang Gerstacker received the Dipl.-Ing. degree in electrical engineering and the Dr.-Ing. and Habilitation degrees from the Friedrich-Alexander University (FAU) Erlangen-Nü rnberg (FAU), Erlangen, Germany, in 1991, 1998, and 2004, respectively. Since 2002, he has been with the Chair of Mobile Communications and Institute for Digital Communications, FAU Erlangen-Nürnberg, where he is currently a Professor. His research interests include the broad areas of digital communications and statistical signal processing, THz communications, 5G and Beyond, and wireless sensor networks. He has conducted various projects with partners from industry.

Dr. Gerstacker was a recipient of several awards, including the Research Award of the German Society for Information Technology (ITG) in 2001, the EEEfCOM Innovation Award in 2003, the Vodafone Innovation Award in 2004, the Best Paper Award of EURASIP Signal Processing in 2006, and the ”Mobile Satellite & Positioning” Track Pa- per Award of VTC2011-Spring. He has been the Technical Program Co-Chair or General Co-Chair of several confer- ences including BlackSeaCom 2014, VTC2013-Fall, ACM NanoCom 2016, and BalkanCom 2019. He is a member of the Executive Editorial Committee of the IEEE TRANS- ACTIONS ON WIRELESS COMMUNICATIONS and an Edi- tor of Computer Networks (Elsevier). He has served as a member of the Editorial Board for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, Physical Communica- tion (Elsevier), and EURASIP Journal on Wireless Commu- nications and Networking, and as a Guest Editor for sev- eral journals and magazines including the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, IEEE COM- MUNICATIONS MAGAZINE, and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.

INDEX OF AUTHORS

International
Telecommunication
Union

Telecommunication
Standardization Bureau (TSB)

Place des Nations
CH-1211 Geneva 20
Switzerland

ISSN: 2616-8375
Published in Switzerland
Geneva, October 2021