

## Survey paper

# Going beyond RF: A survey on how AI-enabled multimodal beamforming will shape the NextG standard

Debashri Roy <sup>a,\*</sup>, Batool Salehi <sup>a</sup>, Stella Banou <sup>a</sup>, Subhramoy Mohanti <sup>a</sup>, Guillem Reus-Muns <sup>a</sup>, Mauro Belgiovine <sup>a</sup>, Prashant Ganesh <sup>b</sup>, Chris Dick <sup>c</sup>, Kaushik Chowdhury <sup>a</sup>

<sup>a</sup> Electrical and Computer Engineering, Northeastern University, Boston, MA 02115, United States of America

<sup>b</sup> Mechanical and Aerospace Engineering, University of Florida REEF, Shalimar, FL 32579, United States of America

<sup>c</sup> Nvidia Inc., United States of America

## ARTICLE INFO

## Keywords:

Beamforming  
Beam selection  
Beam search  
mmWave  
Multimodal  
non-RF data  
Fusion  
5G  
NextG

## ABSTRACT

Incorporating artificial intelligence and machine learning (AI/ML) methods within the 5G wireless standard promises autonomous network behavior and ultra-low-latency reconfiguration. However, the effort so far has purely focused on learning from radio frequency (RF) signals. Future standards and next-generation (nextG) networks beyond 5G will have two significant evolution over the state-of-the-art 5G implementations: (i) massive number of antenna elements, scaling up to hundreds-to-thousands in number, and (ii) inclusion of AI/ML in the critical path of the network reconfiguration process that can access sensor feeds from a variety of RF and non-RF sources. While the former allows unprecedented flexibility in ‘beamforming’, where signals combine constructively at a target receiver, the latter enables the network with enhanced situation awareness not captured by a single and isolated data modality. This survey presents a thorough analysis of the different approaches used for beamforming today, focusing on mmWave bands, and then proceeds to make a compelling case for considering non-RF sensor data from multiple modalities, such as LiDAR, Radar, and GPS for increasing beamforming directional accuracy and reducing processing time. This so called idea of *multimodal beamforming* will require deep learning based fusion techniques, which will serve to augment the current RF-only and classical signal processing methods that do not scale well for massive antenna arrays. The survey describes relevant deep learning architectures for multimodal beamforming, identifies computational challenges and the role of edge computing in this process, dataset generation tools, and finally, lists open research challenges that the community should tackle to realize this transformative vision of the future of beamforming.

## 1. Introduction

Today’s ultra-connected world is demanding high bandwidths, ultra-low latency, and autonomous network reconfiguration to accommodate new applications, heterogeneous devices and shared spectrum use. The number of users is also increasing at unprecedented levels, with predictions of the number of networked devices exceeding 3× the global population by 2023 [1]. To serve bandwidth-hungry application needs, the expected maximum 5G data rate is now revised to be 13× faster in 2023, a significant revision from earlier estimations made only a few years ago in 2018 [1]. Many exciting applications will leverage such high capacity wireless networks, such as relaying high-resolution three dimensional (3D) graphical content, virtual or augmented reality (VR/AR) streams [2], vehicle-to-everything (V2X) links leading towards autonomous cars [3,4], among others.

A key underlying technology that is essential for all of the above is *transmit beamforming*, where signals from multiple antenna elements combine constructively at the receiver. Consider a multi-antenna radio, with each of these antenna elements having a specific directional radiation pattern, referred as a *beam*. The beams from transmitter and receiver antennas are steered to initiate communication via *beamforming* [5]. The communication link is then established through the periodic *beam sweeping* and *beam selection* [6]. Beamforming increases the signal strength at the receiver, which in turn raises the capacity limit, mitigates interference by avoiding undesirable signals at neighboring receivers, and combats the effect of pronounced path loss at high frequencies. Thus, beamforming is considered as a critical component

\* Corresponding author.

E-mail addresses: [droy@ece.neu.edu](mailto:droy@ece.neu.edu) (D. Roy), [bsalehi@ece.neu.edu](mailto:bsalehi@ece.neu.edu) (B. Salehi), [sbanou@ece.neu.edu](mailto:sbanou@ece.neu.edu) (S. Banou), [smohanti@ece.neu.edu](mailto:smohanti@ece.neu.edu) (S. Mohanti), [greusmuns@ece.neu.edu](mailto:greusmuns@ece.neu.edu) (G. Reus-Muns), [mbelgiovine@ece.neu.edu](mailto:mbelgiovine@ece.neu.edu) (M. Belgiovine), [prashant.ganesh@ufl.edu](mailto:prashant.ganesh@ufl.edu) (P. Ganesh), [cdick@nvidia.com](mailto:cdick@nvidia.com) (C. Dick), [krc@ece.neu.edu](mailto:krc@ece.neu.edu) (K. Chowdhury).

<https://doi.org/10.1016/j.comnet.2023.109729>

Received 28 September 2021; Received in revised form 6 February 2023; Accepted 20 March 2023

Available online 23 March 2023

1389-1286/© 2023 Elsevier B.V. All rights reserved.

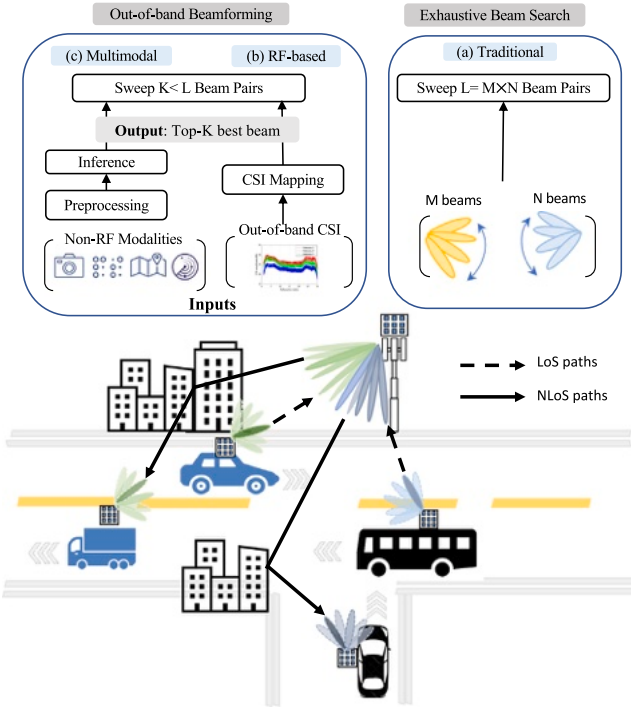


Fig. 1. An overview of different approaches for beamforming in an example scenario involving a mmWave vehicular network. The strategies of beamforming between the roadside base station (BS) and the vehicle is categorized into three types: (a) *traditional exhaustive beam search* that sweeps through all possible mmWave beam combinations between the receiver and transmitter, (b) *RF-based out-of-band beamforming* that uses channel state information (CSI) measurements from lower frequencies to restrict mmWave beam search space, (c) *multimodal beamforming* that uses non-RF sensor modalities (image, LiDAR, GPS, radar) to predict the best possible beams from the situational information.

of all modern WiFi [7] standards and is steadily being integrated into 5G [8].

Our survey is motivated by this observation, and we strive to answer the following two questions: (i) are there fundamental limitations of traditional RF-only beamforming technology that will impact future standards evolution, and (ii) how can new data types (beyond RF) be harnessed in the future, and, given the possible information explosion by acquiring such multimodal sensor feeds, can they be analyzed through emerging machine learning methods to guide real-time beamforming decisions? To ensure a focused discussion, we emphasize use-cases that will shape the future standards beyond 5G (henceforth referred to as *NextG*), namely, beamforming scenarios that combine very large number of antenna elements and mobility. As an indicative example of a mmWave vehicular network that we cover in this survey, Fig. 1 shows moving vehicles beamforming towards a static base station by combining data from RF and non-RF modalities, and then using ML to identify a smaller set of beam-pairs for further optimization, instead of an exhaustive search. RF-based input data needed for mmWave beamforming typically include but are not limited to:

- Channel state information, which includes path loss, shadowing, fading, and multi-path effects, and
- Receiver feedback on the quality of the received signal, which can be leveraged to adjust the beamforming weights and beam selection for optimal performance.

The use of non-RF data as input for mmWave beamforming is geared towards improving the beamforming performance by providing additional information about the environment and system conditions, which are otherwise impossible to decipher from RF-only sources. Non-RF-based input from cameras, LiDAR, and other sensors can be used

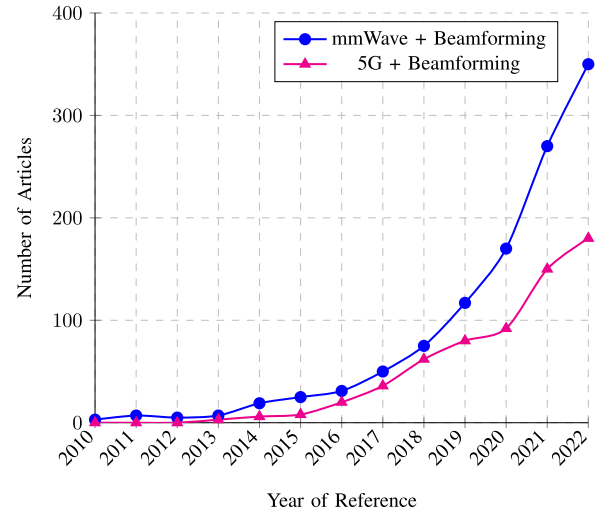


Fig. 2. Number of articles referencing beamforming in 5G for mmWave.

in conjunction with RF-based sources in mmWave beamforming to decipher the:

- User location and movement information, which can be used to optimize the beamforming weights.
- Comprehensive situational awareness, imparting the full knowledge of the environment, including obstacles and potential reflections that can impact the performance of mmWave communications.
- Information about device orientation and position which can be used to improve the directionality of beamforming.
- Information about the location of network access points, which can be used to optimize the beamforming in multi-user scenarios.

Overall, in all types of beamforming methods, the use of multimodal sensor data can lead to increased robustness by providing redundant sources of information and allowing for an enhanced ability to respond to changes in the environment in real-time. We begin our discussion by highlighting the need for beamforming with massive number of antennas and the use of AI/ML in beamforming communication systems.

**• Need for Beamforming in NextG Standards:** The 5G New Radio (5G-NR) standard provisions for use of both sub-6 GHz as well as millimeter wave (mmWave) frequency bands from 24.25 GHz to 52.6 GHz [8]. The sub-6 GHz band is already congested, and this problem worsens when a large data transfer needs to occur at short contact times, typically seen in mobile environments with few antennas [6]. While mmWave-band transmission increases capacity using wider bandwidth (up to 2 GHz) [9], it also suffers from severe attenuation and penetration loss [6]. Phased-array antennas [10] address the attenuation problem by leveraging the highly directional gain of the antenna elements, thereby focusing radiated RF energy into beams. This capability is enhanced in higher frequencies given the dense packing of antenna elements, i.e., higher order phased arrays are possible with proportional increase in the number of beams. While theoretically hundreds of antenna elements can be packed in a 1cm × 1cm area for mmWave band operation, the bottleneck lies in the complexity of processing methods and the computational resource available to properly configure the beams. Even though it is economically feasible to create large phased arrays, scaling beyond 8-12 antennas while supporting real-time operation in small form factor wireless devices still remains an open challenge. Thus, there is need to re-visit existing approaches to beamforming to potentially scale up to thousands of antenna elements, as is being envisaged in NextG standards [11].

• **Motivation for using AI-enabled Beamforming:** Traditional beamforming techniques are restricted by linear operations and have stringent requirements for accurate channel state information and are thus susceptible to slight variations in channel conditions [12] [13] [14]. The integration of AI/ML in traditional beamforming scenarios can address these limitations by incorporating non-linear operations and allowing ML algorithms to efficiently model complex channel dynamics. This approach gives ML-based beamforming models to better adapt to rapidly changing channel conditions and determine optimal beamforming strategies in real time, leading to the realization of more accurate and efficient communication systems.

Artificial intelligence and machine learning (AI/ML) based algorithms have been effectively demonstrated to outperform classical approaches in wireless-centric tasks of modulation recognition [15], RF fingerprinting [16], and rogue transmitter detection [17], etc. The use of AI-enabled algorithms to solve the above-mentioned beamforming in nextG networks is still in a nascent stage. The general approach so far in using ML involves RF channel estimation followed by channel equalization by using different neural network-based architectures that accept a stream of in/quadrature phase (I/Q) samples collected by the receiver. We believe there is a vast untapped potential for AI-enabled techniques for extracting relevant information using different types of modalities, for e.g., images can recognize the location of the target BS and this can rapidly reduce the number of candidate beams to be explored. We refer to this emerging research trend in the domain of out-of-band beamforming as *multimodal beamforming*.

• **Scope of this Survey:** The statistics presented in Fig. 2, comprise of the number of articles (including patents), from Google Scholar search results, that have referenced the terms *beamforming in 5G* and *beamforming in mmWave*. We believe this survey will serve the wireless research community working on beamforming in high frequency band (30-300 GHz), as in these frequencies, beamforming lies on the critical path to combat signal attenuation. We introduce and analyze the notion of multimodal beamforming for mmWave frequencies by recognizing the existing interest in the intersection of MIMO systems, wireless AI/ML and the NextG bandwidth needs. Furthermore, we emphasise the vehicular scenario shown in Fig. 1, as it poses challenges caused by mobility that cannot be addressed in feasible time-scales through legacy methods for such large beamforming antenna arrays. As evidence of community interest on this general theme, we see a spike in citations (178 citations within 4 years) for the publicly available dataset called Raymobtime [18]), which contains multimodal non-RF sensor data along with the corresponding RF ground truths for the purpose of mmWave beamforming in a V2X environment.

While we strive to produce a comprehensive survey on this subject matter, we skip the reviews on the basics of mmWave channel models, mMIMO, different beamforming system models and techniques, as there exist plethora of survey literature focusing on these fundamentals, and is out-of-scope considering our focus area. For example, the promise of mmWave communication in 5G is extensively reviewed in [19], the use of mmWave band for vehicular communication is surveyed in [20], applications of mMIMO are surveyed in [21] and [22], detailed analysis of general RF-only beamforming in indoor and outdoor mmWave communications can be found in [23]. RF-only beamforming can have digital and analog beamforming, as well as hybrid approaches that combine the two. Related models and system architectures that contrast these three approaches are described in [24] and [25]. A flow-graph summarizing the existing surveys related to the “beamforming in 5G/NextG” systems is shown in Fig. 3, and we explore each of these topics in their relevant sections later in this paper. We broadly categorize the trend of existing surveys on that topic in three groups: beamforming techniques for 5G, hybrid beamforming, and out-of-band beamforming; where the first two categories are related to the traditional beamforming process, while the last one is aligned towards out-of-the-box solutions. In this regard, the purpose of this survey is to identify the shortcomings in the traditional beamforming methods

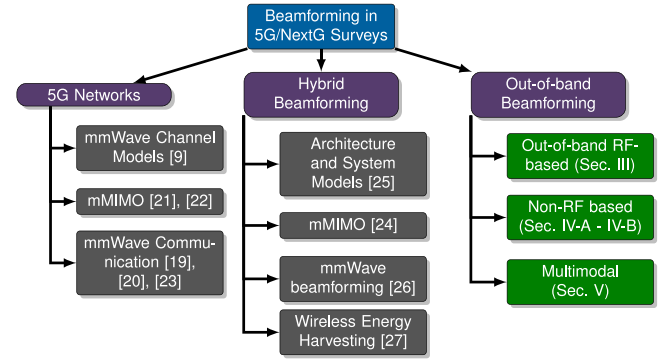


Fig. 3. Existing surveys on different areas of beamforming for 5G and beyond. This survey mainly focuses on use of *out-of-band* beamforming in recent literature for the 5G and NextG networks. We describe different categories of out-of-band beamforming in the corresponding sections in the rest of the paper.

and identify the advantages of using non-RF modalities to facilitate the beamforming process, considering nextG communications. Ultimately, we make a case for expanding the research focus towards incorporating such non-RF sensor modalities in combination with AI/ML, as a feasible pathway for NextG networks.

• **Organization of this Survey:** The remainder of this article is organized as follows. High level differences between different traditional and non-RF based beamforming techniques for nextG networks are described in Section 2 with a comprehensive review of published surveys in related areas of beamforming. The use of out-of-band RF, single, and multimodal non-RF data for beamforming are presented in Section 3, Section 4, and Section 5, respectively. We discuss different emerging research trends using multimodal beamforming in Section 6. The conclusions are drawn in the last section.

## 2. Background on beamforming techniques and related surveys

In this section, we first analyze the state-of-the-art traditional beamforming techniques and its shortcomings. We then explore the current research on non-RF based beamforming to motivate our intent of using these methodologies to address those shortcomings of traditional RF-only beamforming.

### 2.1. Traditional beamforming

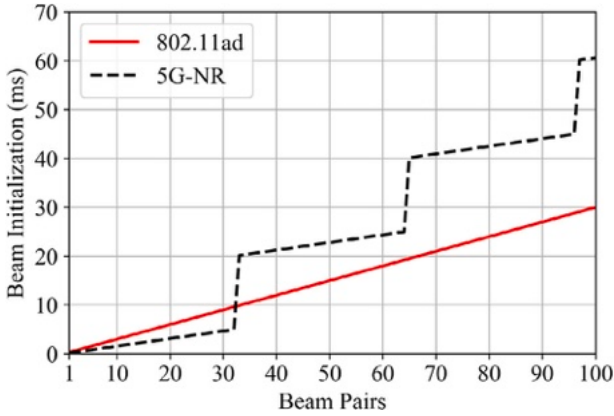
Existing RF-based beamforming approaches (analog, digital, hybrid) have their unique advantages, and are applicable in specific scenarios. Indeed, the 5G-NR standard supports all three types of beamforming in the time domain [23].

A brief comparison study for these approaches is presented in Table 1. Digital beamforming improves the spectral efficiency (SE) of a MIMO system by simultaneously transmitting data to multiple users. However, it needs a distinct RF chain per antenna, making it less cost-effective for higher order of antenna elements. This is one core reason why there are few off-the-shelf mmWave radios [26] which support digital beamforming even with low order (1×4) of antenna elements. Unlike its digital counterpart, analog beamforming creates the beam using one element per set of antenna. Once the best beam, among all possible combinations of beam-pairs is identified, it is activated to mitigate the impact of high pathloss in mmWave band. This is why most of the off-the-shelf mmWave devices [27–29] support only analog beamforming. Also, analog beamforming is considered mandatory in 5G-NR [30] for mmWave communication.

Hybrid beamforming, on the other hand, is a combination of analog and digital beamforming. The idea of hybrid beamforming revolves around trading-off the hardware cost for the overhead of time involved

**Table 1**  
Comparison of different beamforming types.

| Metrics                 | Analog        | Digital                | Hybrid                 |
|-------------------------|---------------|------------------------|------------------------|
| Degree of Freedom       | Limited       | Highest                | High                   |
| Implementation          | Phase Shifter | ADC/DAC, mixers        | Everything             |
| Architecture            | Simple        | Complex                | Complex                |
| Baseband Chains         | Less          | Highest                | High                   |
| Complexity              | Less complex  | Complex                | Complex                |
| Power Consumption       | Less          | Highest                | High                   |
| Cost                    | Less          | Highest                | High                   |
| Inter-user Interference | High          | Lowest                 | Low                    |
| MIMO Support            | No            | Yes                    | Yes                    |
| Flexibility             | Fixed delay   | Flexible weight vector | Flexible weight vector |



**Fig. 4.** The average beam initialization overhead of IEEE 802.11ad and 5G-NR standards with respect to different beam search spaces. The beamforming time increases with increasing number of beam pairs.

in beam selection. Here, a subset of antennas is connected to a particular RF chain, as opposed to having individual RF chains for each antenna element in digital beamforming. Even though hybrid beamforming promises faster communication with higher order antenna elements, this is still an area of ongoing research [31]. Additionally, for hybrid beamforming, the continuous *beam management* technique in a mobile environment involves periodic overhead [30]. Here, beam selection is done after the measurement of reference signals (RS) received in a specific direction by manipulating the beamforming weights applied across different antenna elements.

## 2.2. Out-of-band beamforming

As discussed before, analog beamforming technique involves time-overhead of beam selection due to exhaustive search among all possible transmitter-receiver (TX-RX) antenna elements. The decision is made based on a combination of RF measurements, such as CSI, SNR etc., in the desired frequency band of communication. This overhead (Fig. 4) is exacerbated in the case of mobile users where the position of user equipments (UEs) are changing continuously, resulting in the exhaustive search being instantiated multiple times within a few seconds. Furthermore, the wireless channel varies  $10\times$  faster at 30 GHz as opposed to 3 GHz, even for the same UE mobility rate. This results in  $10\times$  more frequent beam sweeping and channel estimation [32]. Thus, we believe that *out-of-band* RF measurements and the use of environmental non-RF data appear to offer an attractive alternative towards minimizing the overhead of exhaustive search. We refer to such approaches as *out-of-band beamforming* techniques. A visual representation of the existing traditional and out-of-band beamforming techniques are given in Fig. 5.

## 2.3. Beamforming surveys on mMIMO for 5G and beyond

The fundamentals of mMIMO and mmWave operation and the applications of mMIMO are comprehensively surveyed in earlier works [21, 22]. The promise of mmWave communication in 5G is extensively reviewed in [19], and the use of mmWave band for vehicular communication is surveyed in [20].

From Fig. 2 we see that the research interest on beamforming in mmWave band and 5G standards are strongly coupled, as the advancements in the former are essential to meet operational requirements for the latter. Additionally, exploration of new spectrum, assigning more bandwidth, carrier aggregation, inter-cell interference mitigation techniques, integration of mMIMO antennas, etc., are all key features that have been extensively covered in [33]. Also, the authors state that providing accessibility, flexibility, and cloud based services through proper modulation and coding scheme (MCS), mmWave and device to device (D2D) communication is the key to realize functional nextG networks. Authors in [34] validate the notion that beamforming has a bigger role to play in mmWave bands, as compared to low frequency bands. Hence, there is great interest in beamforming optimization in mmWave bands for nextG standards.

For sake of completeness, we mention the surveys that describe beamforming advancements tailored for sub-1 GHz, sub-6 GHz as well as sub-30 GHz 5G bands. Authors in [35] focus on the frequency allocation, beamforming techniques and custom-designed integrated circuits for those specific bands. Kutty et al. capture the evolution of different beamforming techniques in the context of mmWave communication [23]. They describe different radio frequency system design and implementation for mmWave beamforming for indoor and outdoor communication scenarios. The authors describe the mmWave propagation characteristics in-terms of path loss and clustered multipath structures, dominant line-of-sight (LoS) component, wideband communication, and 3D spatio-temporal modeling. They also illustrate different phased array antenna architectures to support MIMO capability in mmWave beamforming. Finally, the authors concur that using hybrid beamforming in the mmWave band for MIMO to minimizing cost and power consumption has great promise.

In a survey on hybrid beamforming for mMIMO, Molisch et al. [24] analyze the trade-offs of using instantaneous or average (second-order) CSI in hybrid beamforming. Here, the authors evaluate current research on various types of hybrid multiple-antenna transceivers and consider how the channel sparsity in the mmWave band can be leveraged for optimizing channel estimation and beam training. However, to get broader aspects of hybrid beamforming, we review an extensive survey by Ahmed et al. in [25], which thoroughly track the progress in this domain till 2017. In this paper the authors present different architectures of hybrid beamforming and the techniques for optimization of phase shifters, DAC/ADC resolutions and antenna configurations. From the system model perspective, they examine eight variations of hybrid beamforming and identify many resource management aspects, particularly in beam management, MAC protocol variants, which can impact the performance of hybrid beamforming.

However, the goal of this survey paper is to describe beamforming techniques that exploit out-of-band RF and non-RF multimodal data for



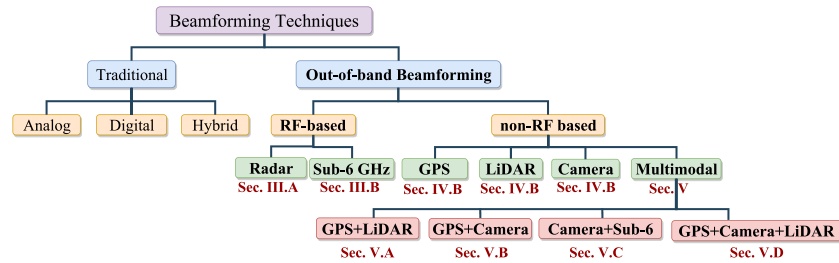


Fig. 5. The general hierarchy of different beamforming techniques, with a focus on out-of-band approaches.

nextG networks. To motivate the case for out-of band RF and multi-modal data, we first identify the limitations of traditional beamforming methods using RF-only data.

#### 2.4. Limitations of traditional RF-only based approaches

The traditional RF-only beamforming approach utilizes one of these two options for mmWave beamforming: (a) estimate the mmWave channel at the receiver, and send this information back to the transmitter for generating the precoding weights [36–39], (b) sweep through the antenna codebook elements of the transmitter and receiver [40,41].

However, the complex method of compressive sensing [42–44] or feedback of channel state information for channel estimation usher in complexity and overhead [45], which must otherwise be kept as low as possible. In order to reduce such overhead of the complicated channel estimation and time-consuming beam-sweeping techniques, multiple out-of-band approaches have been explored in the recent literature, with the aim of achieving low overhead. These beamforming techniques can be broadly categorized into (a) *RF-based* and (b) *non-RF based*, with their different sub-categories illustrated in Fig. 5. In the next sections, we explore in detail each of these categories

### 3. Out-of-band RF based beamforming

The main idea behind leveraging out-of-band RF frequencies during beamforming is to incorporate the *cross channel correlation* at mmWave bands with lower frequencies (2.4 GHz, radar bands, etc.). Such cross correlation is then utilized to reduce the beam search space by establishing a mapping between the channel measurements in the mmWave bands with lower frequencies (see Fig. 6). Although the propagation characteristics in mmWave are different from lower frequencies, recent research reveals that the main direction of arrivals (DoAs) are comparable. Hence, the CSI at lower frequencies can be used to restrict the beam search space and avoid time-intensive exhaustive search, as proposed in the IEEE 802.11ad standard [46]. This is relevant as mmWave systems are very likely to be deployed in conjunction with lower frequency systems, where mmWave access points (APs) are envisioned to be paired with lower frequency APs that provide wide area control signalling and coordination. Moreover, multi-band communication is one of the proposed solutions for providing high throughput communication systems with high reliability, thus reinforcing the interest in taking advantage of such systems in the near future [47]. Among the RF based out-of-band beamforming techniques, the use of radar signals and utilizing sub-6 GHz frequencies for mmWave beamforming have shown promising results.

#### 3.1. Radar

For a vehicle to infrastructure (V2I) hybrid MIMO scenario, González-Prelcic et al. [48], derives the channel information from the infrastructure mounted radar that is used to obtain precoders/combiners at the vehicle and the infrastructure. The radar sensor operates at 76.5 GHz, which is close to the mmWave communication band at 65 GHz. Taking

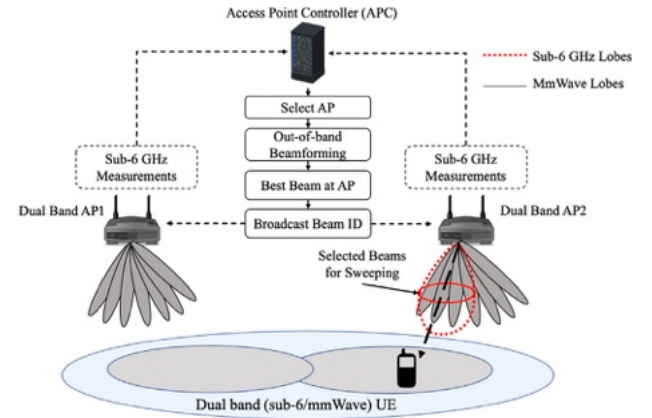


Fig. 6. The AP exploits the measurements at lower frequencies, sub-6 GHz in this example, to propose a set of most likely beam pairs.

advantage of this close proximity of the operating frequencies, the computed covariance of the received signal at the radar is applied as an estimation of the covariance of the communication signal in the mmWave band. The authors then argue that the optimum combiner is the dominant eigenvector of the covariance matrix of the received signal. Similarly, in the proposed scheme by Ali et al. [49], a passive radar at the road side unit (RSU) taps the radar signals transmitted by vehicle mounted automotive radars. In comparison to the prior works, the authors propose a simplified RSU based radar receiver that does not require the transmitted waveform as a reference for covariance estimation in [55]. To use the acquired radar information for mmWave beam initialization, a metric is defined that correlates the spatial information provided by the radar sensor and spatial characteristics of mmWave channel. This metric is then used to assess the accuracy of the angular estimation. Reus et al. [50] leverage the PHY layer IEEE 802.11ad frames to perform both radar operations and conventional communications using the standard compliant TX/RX chain. In this case, the radar is employed to estimate the location of vehicles, which is then used to select the optimal mmWave beam. Similarly, Demirhan et al. [56] use Radar at 77GHz and deep learning to perform beam selection and validate it on the real-world data.

#### 3.2. Sub-6 GHz

Nitsche et al. [51] propose a blind beam steering (BBS) system which couples mmWave with legacy 2.4/5 GHz bands. Upon a beam training request, the proposed method first performs out-of-band direction inference to calculate angular profiles by broadcasting passively overheard frames at the legacy sub-6 GHz band. The LoS paths in all profiles remain nearly static, and appear as peaks at the same angle. However, the peaks resulting from reflections vary among profiles. Thus, using profile history aggregation, the alternating reflection peaks are flattened and the remaining strongest peak is estimated to correspond to the direct path. Given the profile history for each device,

**Table 2**

Survey of literature on out-of-band RF based beamforming.

| Paper                        | Year | Transmission frequency | Dual band frequency | Approach                          | Validation | Multiple AP | Evaluation metric                         |
|------------------------------|------|------------------------|---------------------|-----------------------------------|------------|-------------|-------------------------------------------|
| González-Prelcic et al. [48] | 2016 | 76.5 GHz               | 65 GHz              | Compressive covariance estimation | Simulation | No          | Sum spectral efficiency                   |
| Ali et al. [49]              | 2020 | 73 GHz                 | 76 GHz              | Compressive covariance estimation | Simulation | No          | Transmission rate                         |
| Reus et al. [50]             | 2019 | 60 GHz                 | 5.89 GHz            | Future location estimation        | Simulation | No          | Beamforming time                          |
| Nitsche et al. [51]          | 2015 | 60 GHz                 | 2.4 GHz             | LoS path detection                | Experiment | No          | Direct path estimation accuracy           |
| Ali et al. [52]              | 2017 | 60 GHz                 | 3.5 GHz             | DoA estimation                    | Simulation | No          | Success percentage in detecting best beam |
| Hashemi et al. [53]          | 2018 | 30 GHz                 | 3 GHz               | DoA estimation                    | Experiment | No          | mmWave throughput                         |
| Mohamed et al. [54]          | 2015 | 60 GHz                 | 5 GHz               | WiFi fingerprinting               | Experiment | Yes         | Average packet delay                      |

a threshold for the peak-to-average ratio is defined to infer the LoS path and to reject the reflected paths. The experimental results depict that BBS successfully detects unobstructed direct path conditions with an accuracy of 96.5% and reduces the IEEE 802.11ad beam training overhead by 81%. Similarly, in [52], the candidate mmWave beams are restricted only to those beams that overlap with the dominant paths at sub-6 GHz band. On the other hand, in [53], the estimated angle of arrival (AoA) on the 3 GHz channel is used to reduce the beam sweeping overhead in mmWave 30 GHz frequency. In particular, they experimentally show that in 94% of LoS conditions, the identified AoA in the 3 GHz band is within  $\pm 10^\circ$  accuracy for the AoA of the mmWave signal. Hence, the authors propose using Multiple Signal Classification (MUSIC) algorithm to estimate the AoA in the sub-6 GHz and running the exhaustive search only for angles in the corresponding direction of the mmWave band, while factoring in the error bound of  $\pm 10^\circ$ . A dual-band MAC protocol is proposed in [54] for coordinated wireless gigabit (WiGig) WLANs (see Fig. 6). In the proposed dual-band MAC protocol operation, the control frames to be shared among the APs are transmitted via the wide coverage sub-6 GHz WiFi band, while the high speed data frames are concurrently transmitted by the WiGig APs in the mmWave band. These control frames coordinate the beam training among the APs, so only one AP performs the beam training at a time, eliminating the probability of packet collisions due to simultaneous beamforming. Also, the link information consisting of the used beam identification (ID), modulation coding scheme (MCS) index and received power, is broadcasted in the sub-6 GHz WiFi frequencies, allowing other APs to effectively exclude those beam IDs that may interfere with the existing data link from their beamforming training beams. Moreover, since the location of a UE can be roughly estimated using WiFi channel information at WiFi frequencies through a process called fingerprinting, the authors propose this WiFi fingerprinting method to estimate the best and bad beam IDs of the WiGig links.

We conclude the discussion on out-of-band RF based beamforming techniques by providing a comprehensive overview of these processes in Table 2. Next, we explore the existing challenges in this area.

### 3.3. Challenges

While out-of-band RF assisted beamforming present promising improvements in beam initialization speed, there are some limitations associated, which we itemize as follows:

- The out-of-band RF channel measurements need to be acquired constantly in order to estimate the channel at the mmWave band. Hence, it requires an integrated protocol for multi-band coexistence that can be challenging in the dense networks.
- An optimal mapping is required between mmWave and out-of-band channel measurements. The mmWave band has unique propagation characteristics that preserves sparsity. In particular, the number of reflections is limited in mmWave band, while in lower frequencies, multiple reflections are normally observed. As a result, translating the DoA for bands that are located far apart from each other can be challenging and error prone.

- RF-based out-of-band beamforming requires simultaneous multi-band channel measurements that increase the complexity of mmWave transceivers. Although future mmWave devices will likely support lower frequencies as well, this feature is not widely deployed in commercial devices yet.
- The existing out-of-band methods do not yet support simultaneous beamforming at both the transmitter and receiver sides, which is required for effective directional transmissions.

After motivating the utility of leveraging various non-RF sensor data for RF tasks, we next map these benefits to the use-case of beamforming in mmWave bands, when higher magnitude of antenna elements (i.e., mMIMO systems) are involved. Additionally, the challenges of using RF-based out-of-band beamforming, described in Section 3.3, suggest the research community needs to explore the space of beyond RF-only solutions (be it traditional or RF-based out-of-band). We explore this direction in the next sections.

## 4. Out-of-band non-RF based beamforming

In mmWave beamforming, the location of the TX-RX pair and potential obstacles are the key factors that directly affect the optimal beam configuration. Out-of-band RF aided beamforming methods estimate the approximate location of TX-RX pair given the AoA in other RF bands, which helps to narrow down the beam search space. Interestingly, the situational state of the environment can also be acquired through data obtained from other sensor devices [57], without occupying limited sub-6GHz RF resources. This motivates the use of non-RF sensor data to speed up the beam initialization process in mmWave band [58]. Unlike the previously discussed out-of-band RF methods, the non-RF based beamforming does not require simultaneous multi-band channel measurements and optimal mapping between mmWave and CSI collected from another band. It is also capable of generating a mutually acceptable decision for both transmitter and receiver.

Typically, non-RF based beamforming utilizes inputs from a number of different sensors such as, GPS (Global Positioning System), camera, and LiDAR (Light Detection and Ranging) etc. This is further aided by the fact that with the wide proliferation of IoT, multiple sensors are embedded in the environment, thus making it feasible to obtain situational information from non-RF sources. As an example, consider the automotive sector with vehicles that have advanced driver-assistance systems (ADAS). Fig. 7 depicts the increase in the market revenue of the various sensors enabling ADAS, as reported by Yole Développement [59]. It is expected that the global market for GPS, radar, cameras and LiDARs will reach \$159.6 billion in 2025. With the easy availability of such multitude of sensors, we need to incorporate methods that leverage the heterogeneous sensor data to extract a rich understanding of the environment.

In LoS scenarios, even though the optimal beam configuration can be estimated using the location of transmitter and receiver, it is not trivial to employ such approaches when encountering irregular radiation patterns, for e.g., when devices have multiple side lobes. The problem becomes more challenging when estimating the strongest reflection from obstacles in non-line-of-sight (NLoS) conditions. Hence,

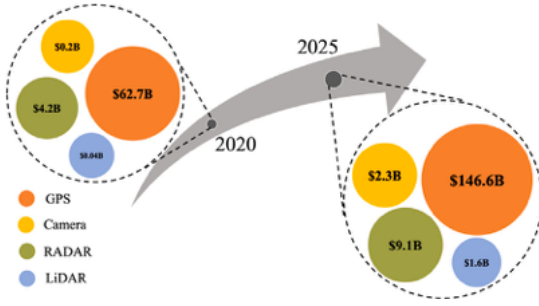


Fig. 7. Market revenue forecast for advanced driver assistance system (ADAS) sensors from the year 2020 to 2025.

a proactive method is required to learn the channel characteristics associated with the observed non-RF sensor modalities on a *case-by-case* basis. Both deterministic and AI-enabled methods are proposed in literature that consider either single sensor modalities or multiple modalities through deep learning. We next go through these state-of-the-art methods, covering different black input sensor acquisition techniques, available datasets, exploitation methods of single and multiple modalities, and identify future research trends.

#### 4.1. Input data acquisition and processing

Choosing the right subset of sensor modalities to accurately capture the environment for detecting potential LoS paths and reflections affecting mmWave frequencies is crucial. The most popular input sensor modalities for mmWave beamforming are presented below and their features are summarized in Table 3.

##### 4.1.1. GPS

This is a popular and widely available satellite-based localization system that generates readings in the decimal degrees (DD) format, where the separation between each line of latitude or longitude (representing 1° difference) is expressed as a float with 5 digit precision. Each measurement results in two numbers that together pinpoints the location on the earth's surface. While localization accuracy in outdoor can be up to 2 m, it drastically decreases in indoor environment because of GPS signal attenuation through walls and structures. It is to be noted that the GPS sensor data refers to the latitude and longitude values generated from the GPS receiver, not the RF signals which are transmitted from the GPS satellites.

##### 4.1.2. Image

Cameras can be used to capture still RGB images of the environment and are commonly used in different applications such as cell phones and surveillance monitoring. Although images allow comprehensive environmental assessment, they are impacted by low-light conditions and obstructions.

##### 4.1.3. LiDAR

The LiDAR (Light Detection And Ranging) sensor generates a 3-D representation of the environment by emitting pulsed laser beams. The distance of each individual object from the origin (i.e., the sensor location) is then calculated based on reflection times [60]. The raw data in LiDAR is defined as the unprocessed data collected by the LiDAR sensor, usually represented as a point cloud representing the position and reflectivity of objects in the environment, along with the time-of-flight information for each measurement. Each point in the cloud represents the distance to the object from where the laser beam reflected back to the LiDAR sensor, along with additional information such as intensity, reflectivity, and wavelength. The raw data from LiDAR sensors can be used for a variety of applications, including mapping, obstacle

Table 3

Sensing technologies for aiding the beamforming.

| Sensing technology  | GPS | Camera | LiDAR |
|---------------------|-----|--------|-------|
| Cost                | Low | Low    | High  |
| Horizontal Range    | –   | High   | High  |
| Resolution          | Low | High   | High  |
| Night Vision        | Yes | No     | Yes   |
| Affected by Weather | No  | Yes    | Yes   |

detection, and autonomous navigation. This additional information can be used in conjunction with user GPS location and RF data for more efficient mmWave beam selection during beamforming [61]. LiDAR can achieve much higher accuracy than image only, but it is expensive and sensitive to weather conditions.

Even if a judicious choice is made on the sensor modality, simply using raw data might fail to provide an accurate prediction. In particular, preprocessing of the raw data can improve the system performance many-folds as we describe later in this paper. Raw observations are not useful unless the role of each device that senses the data is specified, i.e. is the data captured from a transmitter, receiver, or a potential obstacle? Each sensor type has its advantages and limitations. For example, GPS-equipped objects can be utilized to track location, but these sensors cannot capture the presence of obstacles. LiDAR can collect the 3D state of the environment by detecting reflections of emitted pulsed laser light from surrounding obstacles, but it fails to track the location of the target transceivers because LiDAR operates in the visible or near-infrared spectrum and does not detect RF signals. Thus, GPS data can be merged with raw LiDAR data in the preprocessing step to mark the coordinates of the target receiver in the collected point clouds. Hence, data-level aggregation methods are one of the commonly used approaches to refine the raw data to be more informative.

Similarly, the preprocessing steps are also beneficial for reducing the data complexity by either discarding the irrelevant information or reducing the dimensionality of the input data. As an example, using a low-pass filter on camera images can reduce the dimensionality of the image by averaging the adjacent pixels while preserving the integrity. ML-based solutions only accept the data arranged in a fixed size, while for some modalities such as LiDAR the number of point clouds is varying on a case-by-case basis, depending on the number of present objects. Hence, preprocessing can account for this issue by transforming the data to a constrained representation without degrading information content. Hence, it is important to design proper preprocessing steps before using the data for inference. It should be noted here that the preprocessing pipeline of each modality must be designed based on the unique properties of each sensor type.

#### 4.2. Beamforming using single non-RF modalities

Next, we present detailed descriptions of different studies and algorithms that use a single non-RF sensor modality. These include either GPS coordinates, camera images or LiDAR point clouds to accelerate the beam selection, and by extension, the beamforming process.

**•GPS Coordinates:** The knowledge of the location of target receiver has been used earlier to address the challenges of cell discovery [62]. The same idea can be used to speed up the beam initialization in mmWave band, which utilizes directional transmission. The authors in [63,64] use the GPS based position of the receiver to estimate the optimum future beam directions. In particular, the proposed algorithms predict the future locations by tracking the mobility profile of the receiver and geometrical features of the environment. However, it should be noted that this approach only works when the LoS path is available. Alternatively, Wang et al. propose a framework for mmWave beam prediction by exploiting the situational awareness [65]. They use the location of all the vehicles in the same scene as features to extend the solution to NLoS scenarios. The simulation scenario consists of small

cars and trucks, any of which can be the target receiver. The authors argue that the vehicle dynamics have the main effect on the optimum beam configuration, since the road side buildings and infrastructures are stationary, and pedestrians are small in size. Hence, a feature vector map  $v = [r, t_1, t_2, c_1, c_2]$  is generated where  $r$  depicts the location of RSU,  $t$  and  $c$  represent the truck and car vehicles. The subscripts 1 and 2 denote the lane index where the vehicle is located and each vector  $(t_i, c_i)$ ,  $i = 1, 2$  includes the location of the corresponding vehicle type in ascending order for the lane  $i$ . Since the ML algorithms accept a fixed size input, the number of trucks/cars on each lane is constrained, and the vehicles which are far away are eliminated. This feature vector is then used to predict the received power for any beam in the codebook, by leveraging ML.

Similarly, Va et al. [66] propose an algorithm where the GPS location of all the vehicles on the road, including the target receiver, is used as input to a statistical analysis-based algorithm, to infer the best beam configuration. The proposed algorithm uses the power loss probability as a metric to estimate the misalignment probability that might occur when non-optimal beams are selected. Power loss probability defines the probability that the received signal power at the receiver will be less than a certain threshold value. Beamforming solutions are designed to maximize the received signal power at the receiver while maintaining the required signal quality and reliability, which can be done through optimization algorithms that minimize the power loss probability subject to constraints on the transmission power and received noise power. By reducing the power loss probability through proper beam selection and beam width selection, the beamforming solution ensures that the transmitted signal will be received with high quality. In this case, a subset of the beam configurations is suggested by the authors to minimize this misalignment probability.

In order to speed up the beam initialization, an online learning algorithm is proposed in [67], which exploits the coarse user location information in vehicular systems. In particular, the problem is modeled as a contextual multi armed bandit (MAB) problem and a lightweight context-aware online learning algorithm, namely fast machine learning (FML) is used to learn from and adapt to the environment. The proposed FML algorithm explores different beams over time while accounting for contextual information (i.e., vehicles' direction of arrival) and adapts the future beams accordingly, in order to account for the system dynamics such as the appearance of blockages and changes in traffic patterns. In comparison, in [68] Aviles et al. first generate a database that captures the propagation characteristics at 28 GHz and the position of UE. Then, given the location of a UE, a hierarchical alignment scheme is proposed, which consults with this database and incorporates the position of the UE for faster beam alignment.

**•Camera Images and Light Sensors:** The cameras are one of the sensing modalities that capture the situational state of the environment with high resolution. With the recent progress in computer vision and deep learning, powerful algorithms are now available that can be used for processing the images in real time for beamforming. A baseline for ViWi-BT dataset is presented in [69] based on gated recurrent units (GRUs) without the images and only the sequence of beam indices. Since GRUs are a type of Recurrent Neural Network (RNN) which are used to capture long-term dependencies in time-series data [70] and can be used to select the beam that has the highest SNR during beamforming. Alrabeiah et al. argue that beam prediction accuracy is expected to improve significantly by leveraging both wireless and visual data [69]. This strategy is proven to be efficient by the authors in [71], where they use LiDAR along with RF data to predict future mmWave beam selection decisions with high accuracy. In [72], Tian et al. propose a framework to predict future beam indices from previously observed beam indices and images. The proposed approach consists of three steps as follows. The first step consists of feature extraction, where ResNet, ResNext and 3D ResNext modules, each proven to have powerful feature-representation abilities, are used to capture 2D and 3D spatiotemporal visual and motion features from the

input time series data in the form of images. In the second step, these features are merged through the use of a Feature Fusion Module (FFM) that comprises two long short-term memory (LSTM) [73] networks for aggregating the features, followed by a crossgating block to make full use of related semantic information between these two features by multiplication and summation. To validate their approach, the authors use ViWi-BT dataset where the first eight pairs of images are used as time series input data to predict the next five future beams.

Similarly, in [74], Xu et al. propose a scheme where the images captured from different perspectives are used to construct a 3D scene that resembles the point cloud data collected by 3D sensors like LiDAR. Then, a CNN with 3D input is designed to predict the future beams to be selected. Results reveal that the proposed 3D scene based beam selection outperforms LiDAR in accuracy, without imposing the huge cost of LiDAR sensor. While the majority of current literature uses synthetic datasets, the authors in [75] deploy a testbed using National Instruments radio at 60 GHz [27] and camera generated images to predict the best beam configuration. Their proposed method consists of two main steps, namely detection and prediction. In the first step, the transmitter and receiver are detected in the image in the form of a bitmap. This step is important to detect the features which are relevant to the task and discard the irrelevant ones, such as static walls, etc. Finally, the bitmaps are fed to another CNN to predict the optimum beam configuration given the historical data from collected dataset. The LiSteer system proposed in [76] steers mmWave beams to mobile devices by re-purposing indicator light emitting diodes (LEDs) on wireless APs to passively track the direction to the AP using light intensity measurements with off-the-shelf light sensors. The proposed approach considers the pseudo-optical properties of mmWave signal, i.e., dominant LoS propagation, to approximate the APs' AoA in the mmWave band. Hence, their approach requires the APs to be equipped with LEDs and to be situated close to the mmWave band antenna. The authors propose using an array of light sensors to combat the in-coherency of light-AoA estimation that also allows steering beams for both 2D and 3D beamforming codebooks. The experimental results demonstrate that LiSteer achieves direction estimates within  $2.5^\circ$  of ground truth on average with beam steering accuracy of more than 97% in tracking mode, without incurring any client beam training or feedback overhead.

**•LiDAR Point Clouds:** Woodford et al. [77] use LiDAR to build a 3D map of the surrounding physical environment and captures the characteristics of the physical materials. The proposed approach uses a customized ray-tracing algorithm that can identify real RF paths in a 3D mesh generated by LiDAR sensors, and reject false reflection paths caused by reconstruction noise. The output of this phase is a pre-computed look-up table to select the best beams for all mmWave links in the environment. It should be noted that the LiDAR sensors are not required during the ordinary operation of the system and are only used in advance to generate the lookup table. The proposed approach can recompute the complete lookup table for the environment within 15 minutes. The authors validate their approach using Azure Kinect LiDAR camera [78] and a commercial 802.11ad radio [79], yielding to 66% reduction in latency and 50% increase in throughput. Similarly, Jian et al. [80] use the LiDAR data to predict future beams in a V2I scenario, without requiring any knowledge about the previous optimal beams. Thus, a sequence of captured LiDAR sensing images are used as input to a recurrent neural network (RNN) to predict the top-K promising beams. The proposed approach is validated on the publicly available DeepSense 6G dataset [81] and the results indicate slightly lower accuracy than a baseline model that has perfect knowledge of the previous optimal beams.

A concise overview of different state-of-the-art beamforming methods while using single sensor data is presented in Table 4.



**Table 4**  
Survey of literature on non-RF data for out-of-band beamforming.

| Paper                 | Year | Frequency | Data type    | Approach      | Validation | Control channel | Evaluation metric             | Dataset publicly available? |
|-----------------------|------|-----------|--------------|---------------|------------|-----------------|-------------------------------|-----------------------------|
| Kim et al. [63]       | 2013 | 60 GHz    | GPS          | Deterministic | Simulation | No              | Achievable capacity           | No                          |
| Va et al. [64]        | 2016 | 60 GHz    | GPS          | Deterministic | Simulation | No              | Outage and average rate ratio | No                          |
| Wang et al. [65]      | 2018 | 5 GHz     | GPS          | Deep learning | Simulation | No              | Alignment probability         | No                          |
| Va et al. [66]        | 2017 | 60 GHz    | GPS          | Deep learning | Simulation | Yes             | Power loss probability        | No                          |
| Sim et al. [67]       | 2018 | 28 GHz    | GPS          | MAB           | Simulation | Yes             | Cumulative RX data            | No                          |
| Aviles et al. [68]    | 2016 | 28 GHz    | GPS          | Deterministic | Simulation | Yes             | CDF of AoA estimation         | No                          |
| Alrabeiah et al. [69] | 2020 | 28 GHz    | Camera       | Deep learning | Simulation | No              | Top-1 accuracy                | Yes                         |
| Tian et al. [72]      | 2020 | 28 GHz    | Camera       | Deep learning | Simulation | No              | Top-1 accuracy                | Yes                         |
| Xu et al. [74]        | 2020 | 60 GHz    | Camera       | Deep learning | Simulation | Yes             | Top-K accuracy                | Yes                         |
| Salehi et al. [75]    | 2020 | 60 GHz    | Camera       | Deep learning | Experiment | No              | Top-1 accuracy                | Yes                         |
| Woodford et al. [77]  | 2021 | 28 GHz    | LiDAR        | Deterministic | Mixture    | No              | Link latency                  | No                          |
| Jiang et al. [80]     | 2022 | 60 GHz    | LiDAR        | Deep Learning | Experiment | No              | Top-K accuracy                | Yes                         |
| Haider et al. [76]    | 2018 | 60 GHz    | Light sensor | Deterministic | Experiment | No              | CDF of AoA estimation         | No                          |

#### 4.3. Available datasets with single sensor

All these deep learning based methods warrant for data collection and training needs for the discussed methodologies. Hence, next we discuss the features of the available public datasets specific to beamforming using non-RF sensor modalities. These datasets enable the research community to explore different aspects of non-RF beamforming without incurring an individual effort of data collection.

##### 4.3.1. ViWi

Alrabeiah et al. proposed a scalable synthetic framework called Vision-Wireless (ViWi) [82]. The scenario of interest is a V2I setting in 28 GHz mmWave band. The first release of this dataset contains four scenarios with different camera distributions (co-located and distributed) and views (blocked and direct). The channel characteristics and images are generated using the Remcom Wireless Insite ray-tracing [83] and Blender [84] software, respectively. For each scenario, a set of images and raw wireless data (signal departure/arrival angles, path gains, and channel impulse responses) are recorded. An extended version of this dataset is named ViWi vision-aided mmWave beam tracking (ViWi-BT) [69], which contains 13 pairs of consecutive beam indices and corresponding street view images. This dataset contains a training set with 281,100 samples, a validation set with 120,468 samples, and a test set with 10,000 samples.

##### 4.3.2. Image-based

This dataset is obtained by Salehi et al. in [75] from a testbed composed of two Sibeam mmWave [27] antenna arrays mounted on sliders enabling horizontal movement. Using the mmWave transceivers from National Instruments, the mutual channel is measured for 13 beam directions at transmitter and receiver (169 beam configurations overall). Two GoPro cameras observe the movements in the environment and are synchronized with the mmWave channel measurements. In the designed scheme, an obstacle blocks the LoS path between the transmitter and receiver and the experiment is repeated for two types of obstacles, wood and cardboard, causing 30dB and 4dB attenuation while blocking the LOS path, respectively.

## 5. Multimodal beamforming

Each of the previously mentioned sensor modalities capture different aspects of the environment, using more than one sensor modality and intelligently fusing these multimodal data can result in more comprehensive understanding of the environment and can consequently enable the undertaking of robust decisions.

• **Benefits of Fusion:** The fusion of multimodal data over the single modalities has multiple advantages, as explained below:

- **Enhanced Data Representation:** For the situational information to be effective during beamforming, it is crucial to differentiate between the transmitter, receiver and obstacles. However, some sensor modalities cannot provide such information by only relying on raw data. In this case, the data from different modalities can be fused together to improve the data representation. As an instance, it is not trivial to locate the receiver within a LiDAR point cloud. In this case, the GPS coordinates can be used to mark the target receiver.
- **Compensate for the Missing Information:** Sometimes the captured data from each sensing modality reflect an aspect of the environment, yet none can provide a complete understanding by it's own. For instance the dimensionality of objects is not reflected in GPS, and the accurate Cartesian coordinates of the target receiver cannot be acquired using LiDAR or image sensors.
- **Improved Accuracy:** Using more than one modality enables a fine grained understanding of the environment which results in more accurate predictions. Hence, fusion reinforces the prediction accuracy by gathering the information from different sensors to make the final decision. In this case, the fusion algorithms can automatically adjust the weights of each modality towards the optimum performance.
- **Robustness to Errors:** Collecting data using sensor devices comes with associated considerations, including the inherent error. Here, the accuracy of measurement is dependent on working with the nominal structure that the device is designed for. For instance, the accuracy of LiDAR sensor degrades with sunlight reflections, while it does not affect the GPS data [85]. Hence, fusion increases prediction robustness in the case of inaccurate or unreliable data.
- **Availability:** In some applications, the sensor does not have to be co-located. Hence, secondary control channels are required to enable the connectivity between the different sensors and the computing unit. However, this control channel is also subjected to saturation and loss. Using more than one modality with fusion helps the system to be robust to such scenarios and it guarantees that the prediction happens when at least one modality is available during the inference.

Below, we give some examples of state-of-the-art multimodal beamforming with different fusion approaches on multiple sensors. The following subsections are named after the multimodal sensor names which are used as input for the beam prediction in the state-of-the-art. Data from each type of sensors is processed through the same way as mentioned in Section 4.1.

### 5.1. GPS and LiDAR

Consider a typical V2I setting, where a static BS wants to establish communication with a target vehicle-mounted receiver. The vehicle is assumed to be equipped with GPS and LiDAR sensors that enable

the vehicle to acquire its location in 3D space and generate a map of the environment, including the location and shape of objects in the area, which can be used to detect nearby obstacles. The LiDAR point cloud data can be formalized by the set  $\mathcal{P} = \{p_1, p_2, \dots, p_n\}$  consisting of 3D points. Each point  $p_i = (x_i, y_i, z_i)$  represents the position of the object in 3D space, where  $x_i$ ,  $y_i$ , and  $z_i$  are the coordinates of the point in the X, Y, and Z dimensions, respectively. This spatiotemporal data together with the GPS coordinates of the user can be used to improve the accuracy of mmWave beamforming by allowing the system to avoid beams that would be blocked by obstacles, and instead select beams that will provide the best signal quality. Also, the precise location information obtained by LiDAR and GPS data fusion can be used to improve the precision of beam steering, ensuring that the beams are directed precisely at the intended target, thereby improving performance and ensuring higher data rates in mmWave communication systems.

In this scenario, Klautau et al. propose a distributed architecture to reduce the mmWave beam selection overhead [86]. Here, the BS constantly broadcasts its position via a low-band control channel. The situational state of the environment is then collected using LiDAR, situated on the vehicle and is aggregated by BS location in the pre-processing pipeline, where a histogram is generated at the beginning to quantize the space. The LiDAR point clouds then lie in the corresponding bin of the histogram, and the location of BS and receiver is also marked with unique indicators. Using the proposed preprocessing step, the measured point clouds are mapped to a ridge represented by a fixed size. Note that the number of point clouds in the raw data varies depending on the number of objects present during the measurement. This refined data representation is then fed as input to a deep CNN to estimate a set of  $K$  most likely candidate beam pairs. The selected beam pairs are then sent to the BS, and beam training is performed to generate the suggested subset to obtain the optimum beam configuration. Similarly, Dias et al. consider a V2I setting and compare the performance of the previously described distributed scheme with two centralized schemes: (i) using a single LiDAR located at the BS, and (ii) fusing LiDAR data from neighboring vehicles at the BS [87]. The LiDAR data is then used for both LoS detection and beam selection for three competing scenarios. The experimental results in this work depict that in LoS, distributed and centralized methods perform closely, while the LiDAR at BS results in lower top- $K$  beam prediction accuracy, because of the limited range of LiDAR. On the other hand, in NLoS scenarios, the distributed scheme outperforms the centralized method, and both are better than LiDAR at BS.

## 5.2. Camera with sub-6 GHz

The possibility of vision-aided wireless communications is evaluated in [88] where a camera at the BS observes the movements in the environment, and snapshots of the environment are paired with sub-6 GHz channels to help overcome the beam selection and blockage prediction overhead. The proposed method models the beam prediction from images as an image classification task. Hence, each user location in the scene is mapped to a class representing the associated beamforming codebook. However, the pure image input may be insufficient for blockage detection since the instances of 'no user' and 'blocked user' are visually the same. Hence, in order to identify blocked users, the images are fused by sub-6 GHz channels to account for the aforementioned challenge.

## 5.3. GPS and camera

Reus-Muns et al. [61] propose a deep learning based fusion framework for beam selection which leverages the GPS location information along with visual data yielding to a low-overhead fast beamforming process under mobility scenarios. An object detection algorithm is employed to filter out background clutter and identify different types of vehicles. Moreover, the authors analyze the impact of the

LoS/NLoS conditions in terms of beam selection accuracy and propose a method to predict mmWave link blockage. Similarly, Charan et al. study multimodal beamforming in a drone communication system where the sensory data includes images of the drone captured at the BS, GPS position, height of flight (distance to ground), and projected distance (horizontal distance to BS) [89]. The evaluations on a real-world dataset depict up to 86.32% top-1 accuracy with 64 pre-defined beams. In [90], a deep neural networks predicts the optimal beam using positional and visual (camera) data with 75% top-1 beam prediction accuracy.

## 5.4. GPS, camera and LiDAR

The sensor suite in a multimodal beamforming system can be extended to include a multitude of modalities. In fact, the more comprehensive the situational information are, i.e., more sensors, the more robust the prediction will be. Salehi et al. [41] consider a V2I scenario in which the vehicle (receiver) is equipped with GPS and LiDAR sensors and a roadside camera also tracks the movements in front of the base station from Raymobtime dataset. The observations indicate that fusion of all three modalities improves the beam selection accuracy by 9.9-43.9% with respect to least and most significant single modalities, GPS and LiDAR, respectively. Moreover, the evaluations reveal 95% improvement in beam selection speed over classical RF-only beam sweeping method. Later, the same setting with three sensor modalities is studied in a federated learning setting in [91]. In a centralized learning method the entire multimodal data is transmitted to a central unit for training which is costly. In the proposed federated learning framework, different vehicles collaboratively learn a shared model by performing a) local training using the sensor data on-board b) sharing the local model with the BS c) federated aggregation at the BS to generate the *global model* d) sending the global model in the downlink and start a new federated iteration with global model weights as initialization. The observations show that the proposed framework provides a 46.04% improvement in accuracy over local learning and a 70.05% improvement in overhead over centralized learning. In a similar work, Gu et al. [92] use the GPS, camera and LiDAR sensors for beam selection but extend the sensor suite to two camera angles and two LiDARS configurations.

The concise overview of different state-of-the-art beamforming methods while using multimodal sensing data is presented in Table 5.

## 5.5. Available datasets with multimodal sensors

To validate the deep learning based state-of-the-art methods of multimodal beamforming, we next discuss various data collection process and published datasets for the discussed methodologies.

### 5.5.1. Raymobtime

The Raymobtime multimodal dataset [18] captures a virtual V2X deployment with high fidelity in the urban canyon region of Rosslyn, Virginia for different traffic patterns. A static roadside BS is placed at a height of 4 meters, alongside moving buses, cars, and trucks. The traffic is generated using the Simulator for Urban Mobility (SUMO) software [95], which allows flexibility in changing the vehicular movement patterns. The image and LiDAR sensor data are collected by Blender, and Blender Sensor Simulation (BlenSor) [96] software, respectively. For a so called scene, the framework designates one active receiver out of three possible vehicle types i.e. car, bus and truck. A python orchestrator invokes each software for each scene and collects synchronized samples of LiDAR point clouds, GPS coordinates and camera images mounted at the BS. The combined channel quality of different beam pairs are also generated using Wireless Insite ray-tracing [83] software. The number of codebook elements for BS and the receiver is 32 and 8, respectively, leading to 256 beam configurations overall.

**Table 5**

Survey of literature on multimodal beamforming using non-RF data.

| Paper                 | Year | Frequency | Data type          | Approach      | Validation | Control channel | Evaluation metric | Dataset publicly available? |
|-----------------------|------|-----------|--------------------|---------------|------------|-----------------|-------------------|-----------------------------|
| Klautau et al. [86]   | 2019 | 60 GHz    | GPS and LiDAR      | Deep learning | Simulation | Yes             | Top-K accuracy    | Yes                         |
| Dias et al. [87]      | 2019 | 60 GHz    | GPS and LiDAR      | Deep learning | Simulation | Yes             | Top-K accuracy    | Yes                         |
| Alrabeiah et al. [88] | 2020 | 28 GHz    | Camera and sub-6   | Deep learning | Simulation | Yes             | Top-K accuracy    | Yes                         |
| Charan et al. [89]    | 2022 | 60 GHz    | Camera and GPS     | Deep Learning | Experiment | Yes             | Top-K accuracy    | Yes                         |
| Charan et al. [90]    | 2022 | 60 GHz    | Camera and GPS     | Deep Learning | Experiment | Yes             | Top-K accuracy    | Yes                         |
| Reus-Muns et al. [61] | 2021 | 60 GHz    | Camera and GPS     | Deep learning | Simulation | Yes             | Top-K accuracy    | Yes                         |
| Salehi et al. [41]    | 2022 | 60 GHz    | GPS, Camera, LiDAR | Deep learning | Simulation | Yes             | Top-K accuracy    | Yes                         |
| Salehi et al. [91]    | 2022 | 60 GHz    | GPS, Camera, LiDAR | Deep learning | Experiment | Yes             | Top-K accuracy    | Yes                         |

**Table 6**

The overview of different datasets available for non-RF based (for both single and multimodal) mmWave beamforming.

| Name         | Modalities                                                         | mmWave devices            | Type       | Source |
|--------------|--------------------------------------------------------------------|---------------------------|------------|--------|
| ViWi         | Camera Image                                                       | Wireless Insite [83]      | Synthetic  | [69]   |
| Image-based  | Camera Image                                                       | Sibeam [27]               | Real-world | [75]   |
| Raymobtime   | Camera Image, LiDAR, GPS                                           | Wireless Insite [83]      | Synthetic  | [18]   |
| FLASH        | Camera Image, 16 channel LiDAR, GPS                                | TP-Link Talon AD7200 [93] | Real-world | [94]   |
| e-FLASH      | Front and Side Camera Images, 16 channel and 64 channel LiDAR, GPS | TP-Link Talon AD7200 [93] | Real-world | [94]   |
| DeepSense 6G | Camera Image, Radar, 32 channel LiDAR, GPS                         | SIVERS frontend [81]      | Real-world | [81]   |

### 5.5.2. FLASH and e-FLASH

The FLASH multimodal dataset is the first real-world deployment of a mmWave V2I system. The testbed includes a 2017 Lincoln MKZ Hybrid autonomous vehicle with GPS, Camera and LiDAR sensor on-board. The GoPro Hero4 camera with a field-of-view (FOV) of 130 degrees is faced towards BS and the Velodyne VLP 16 LiDAR is mounted on top of the vehicle. The sensors are connected to a central computer with Robot Operating System (ROS) to store the recordings with their timestamps. On the other hand, the TP-Link Talon AD7200 tri-band routers, with Qualcomm QCA9500 IEEE 802.11ad Wi-Fi chips, are used as the BS and Rx at the 60 GHz frequency [93]. The default codebook includes sector IDs from 1 to 31 and 61-63 for a total of 34 sectors; the sectors with IDs of 32 to 60 are undefined. The multimodal sensor data and RF grand truths are captured at the following rates: 1 Hz for GPS, 30 frames per second (fps) for the camera, 10 Hz for LiDAR, and 1-1.5 Hz for the RF ground-truth. Later, Gu et al., extend the sensor suite in FLASH with two additional sensors. An extra GoPro Hero9 camera is faced towards the road, unlike FLASH where the camera is faced towards the side of the road. Moreover, a Velodyne VLP 64 LiDAR is included to capture the point clouds with higher resolution of 64 channels than the 16-channel LiDAR used in FLASH. The FLASH and e-FLASH dataset include different scenarios such as LoS, NLoS with either pedestrian, static or moving car, and both are released to community and publicly available in [94].

### 5.5.3. DeepSense 6G

Alkhateeb et al. explored a progressive set of 5 testbeds for real-world mmWave band deployments. In the most advanced setting, a multimodal dataset in a real-world multimodal V2I setting is released [81]. While in FLASH dataset all the sensors are deployed on the vehicle, in DeepSense only GPS is available at vehicle (sampling frequency of 10 Hz). The remaining sensors including camera, RADAR, and LiDAR are deployed at the roadside base station. The camera (ZED2 from StereoLabs) covers up to 110 degrees of the field of view (FOV), recording 30 frames per second. On the other hand, the Frequency Modulated CW (FMCW) RADAR with 3 Tx and 4 Rx antennas records the complex I/Q RADAR measurements at frequency range 76-81 GHz. Finally, an Ouster LiDAR with range of 40m and resolution of up to 3cm records the point clouds [56]. The mmWave radio in DeepSense includes a 16-element Uniform Linear Array (ULA) phased array from

SIVERS semiconductors. The radios operate at 60 GHz with 64 pre-defined codebooks that uniformly scan 90° field-of-view. The released dataset for V2I scenario includes up to 18,667 samples.

The overview of different publicly available datasets for both single and multiple modalities for beamforming is presented in Table 6.

### 5.6. Challenges

While multimodal learning is an extremely interesting research field, there are some challenges that need to be addressed. First, in order to exploit more than one modality, the synchronized information of all modalities must be present during inference. This requires a precise network controller and back-channel to enable connectivity among different modules while accounting for privacy concerns. Second, the fusion scheme needs to be designed such that the different modalities result in a reinforced prediction. The fusion model can be as simple as a linear transformation, such as summation or multiplication. However, learning the relation between different modalities might require non-linear transformations such as deep learning on custom-made neural networks. In that regard, exploring different novel fusion techniques using different non-linear transformation is widely researched area of current state-of-the-art.

## 6. Emerging research frontiers using multimodal beamforming

In this section, we present selected research frontiers, where the AI-enabled multimodal beamforming techniques can make transformative difference.

### 6.1. V2X networking

The Vehicle-to-Everything (V2X) market is projected to reach \$13 Billion by 2028 [97]. V2X will enable communication among vehicles as well as between vehicles and networks, infrastructure and pedestrians, aiming to improve traffic efficiency, road safety and individual vehicle energy efficiency [98]. V2X connectivity is also essential for the advancement of autonomous driving. Traffic efficiency improves by monitoring congested areas and providing alternative routes, while maintaining road safety by monitoring speed and identifying risky drivers. At the same time, V2X networks can improve energy efficiency

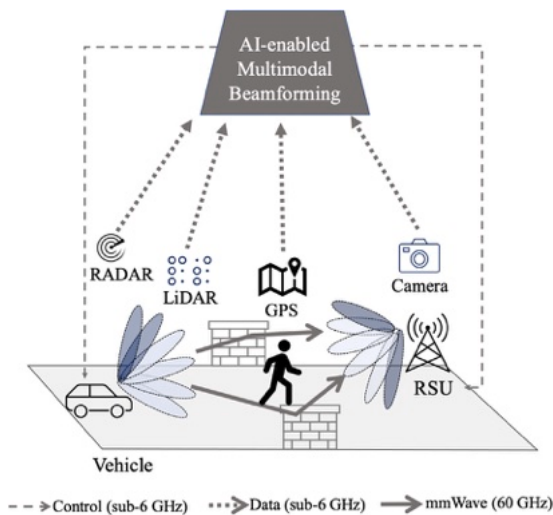


Fig. 8. Pedestrian detection using multimodal beamforming in V2X environment. Different sensor information (LiDAR, RADAR, Camera, and GPS) are leveraged to perform the beamforming between vehicle and road side unit (RSU), as well as pedestrian detection between them.

by making vehicles more intelligent, choosing journeys with lower carbon emissions.

Different beamforming techniques have direct impact on the performance of 5G-V2X networks [99]. In [100], Lee *et al.* presents an object detection algorithm by fusing visual and LiDAR data to form 3D images of the vehicle surroundings. Combining these two concepts, we envision the application area of using multimodal beamforming for V2X architecture will extend from fast and reliable communication to object detection in urban scenarios. As mentioned in Section 4.1, availability of different types of sensor data forms the backbone of V2X communication. The multimodal beamforming using these sensors can be leveraged to provide low latency V2X communication. The knowledge of such selected beams at each specific location can then be leveraged to detect objects or pedestrian using AI-enabled algorithms. An example use-case of pedestrian detection via beamforming is depicted in Fig. 8.

## 6.2. UAV communication

Unmanned aerial vehicles (UAVs) are used extensively in military, scientific and civil applications. They can be used for capturing data, monitoring non-accessible areas and developing high-throughput wireless communication infrastructure. Networks of UAVs, known as flying ad-hoc networks (FANETs) have sparked great interest in academia, industry and government due to their flexibility, low cost and wide range of applications: disaster management, relaying networks, agricultural processes and many more [101]. For all those applications, high-speed low-latency wireless communication is essential between UAVs as well as from UAVs to ground entities (UAV-Ground).

Images captured by flying UAVs may need to be distributed to ground nodes, while data from the ground terminals is required by the UAVs for channel allocation and routing [102]. Distributed beamforming is an important enabler for leveraging high throughput and long range communications through flying UAVs, given their high probability of LoS links due to their altitude. Drawbacks in these scenarios such as inaccurate GPS signals, unpredictable UAV hovering, etc., create the need for accurate transmission synchronization between multiple UAVs through external sensor data input [103], in order to realize a practical distributed beamforming implementation for multi-UAV to ground [104] and UAV-UAV communications. [105] and [106] highlight the use of mmWave links for UAV-UAV and UAV-ground

communication. Ultra-fast UAV communication is essential for wireless infrastructure drones (WIDs). To improve the need for faster and more reliable communication in both the above cases, beamforming in mmWave can be combined with AI-enabled techniques. With the introduction of camera images and other non-RF multimodal data, such as GPS, beamforming in UAV communication can be enhanced with multimodal beamforming to provide higher throughput, robustness, coverage and delay metrics.

## 6.3. Multi-agent robotics

Autonomous agents are increasingly used in a variety of applications like mining [107], agriculture [108], military [109], aerospace [110] and medicine [111] to name a few. Many system entities need to collectively coordinate with each other to make decisions online and collaboratively in this paradigm. Some examples of multi-agent robots are simultaneous localization and mapping (SLAM) [112–114], warehouse robotics [115–117], surgical robotics [118,119], autonomous driving [120–122], and agricultural robotics [123–125] etc. In these applications, each of the agents in a multi-agent system may be equipped with sensors like LiDAR, RGB and infra-red (IR) cameras, and GPS receiver etc., which enable them to function autonomously. Many applications rely on agents being able to communicate locally with other agents. For such applications and for real-time collaboration, high-speed communication is needed for sharing sensor information, decisions and actions. To support such large data-rate requirements, industries with highly automated process flows are pursuing high bandwidth communication links, including access to mmWave bands [126]. Multimodal beamforming using non-RF data can be an interesting approach to facilitate the faster communication between such autonomous entities using the integrated sensors within them.

## 6.4. Terahertz communication

Early works that prove the feasibility of exploiting the THz frequency bands (0.3 THz to 10 THz) point towards an upcoming paradigm shift in the way wireless spectrum will be used. THz-band links bridge the gap between radio and optical frequency ranges, which may be game-changing for nextG wireless networks [127] by enabling transfer rates of 10Gb/s [128]. However, highly directional and fine-grained beams in the phased array antennas, which are essential to support the THz communication, come with their own challenges. Additionally, the beam search space increases with the increasing frequencies. Hence, there is urgent need to exploit out-of-the-box approaches, such as AI-enabled CSI estimation techniques, to decouple the number of antenna elements from the beamforming time overhead [39]. We believe the idea of multimodal beamforming can also be extended to THz communication to reduce the exploding search space of antenna codebook elements by leveraging the environmental multimodal data.

## 6.5. Virtual presence

Since the start of the COVID-19 pandemic, we have quickly transitioned to using virtual communications platforms to aid in wellness and safety. However, platforms like Zoom and Teams can only do so much with respect to quality of user experience. Most of these platforms are still limited by the on-screen presence. This is where the recent development of eXtended Reality (XR) can make a difference by opening up the possibility of transforming *on-screen presence* to a *virtual presence*. The concept of holographic representation can emulate physical presence for meeting, gaming, or collaborating with others. Such virtual presence will support mobility in group presentation or multi-player gaming situations. XR technologies will require multi-Gbps data-rates that may saturate a sub-mmWave band within seconds. Even the still-evolving 5G standard is not capable of supporting these data transfer rates. The standardization of ultra-fast beamforming in mmWave communication



is integral for NextG standards [11]. The concept of using multimodal non-RF data in such applications is promising in this regard. The rich properties of XR or holographic images can be exploited for situational awareness to aid in the beamforming in high frequencies, where the codebook search space is generally too large to compute optimally in real time via exhaustive searching [129].

## 6.6. Hybrid beamforming

mmMIMO communications in hybrid transceivers is realized by a combination of high dimensional analog phase shifters and power amplifiers with lower-dimensional digital signal processing units [25]. For fully connected hybrid transceivers, the situational states through the non-RF modalities can be leveraged to select multiple phase shifters (multi-label prediction), which can be inferred to derive the best RF chains and aid in even-faster beamforming. Multimodal beamforming can be applied per RF chain to select the best phase shifter, and this will enable the parallel inference of all the RF chains at the same time. Hence, the use of multimodal data has huge potential for improving the emerging hybrid beamforming technique and it will allow seamless scaling to make it suitable for NextG networks.

## 7. Conclusions

This paper provides a comprehensive survey of AI-enabled beamforming techniques using out-of-band and multimodal data for mmWave communication in NextG networks. While the previous surveys on beamforming [23–25] are focused more on analyzing and using mmWave channel characteristics, or channel state information for beamforming in massive MIMO leveraging the complicated hybrid beamforming process; our survey reviews recent trends in the literature that adopt an out-of-box approach for solving the same problem. We discuss the state-of-the-art in research trends, application areas, and open challenges of this exciting and emerging paradigm of multimodal sensor data-enabled beamforming. We identify those open research challenges to motivate future research and well as indicate the potential transformative impact of this area on different wireless applications.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

No data was used for the research described in the article.

## Acknowledgment

The authors gratefully acknowledge the funding from the US National Science Foundation (grant CNS-1923789).

## References

- [1] Cisco, Cisco annual internet report (2018–2023) white paper, 2020, [Online]. Available: <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>.
- [2] W. Haerick, M. Gupta, 5G and the factories of the future, 2015, pp. 1–31, [Online]. Available: <https://5g-ppp.eu/wp-content/uploads/2014/02/5G-PPP-White-Paper-on-Factories-of-the-Future-Vertical-Sector.pdf>.
- [3] Huawei, 5G: A technology vision, 2014, [https://www.huawei.com/mediafiles/CORPORATE/PDF/Magazine/WinWin/HW\\_329327.pdf](https://www.huawei.com/mediafiles/CORPORATE/PDF/Magazine/WinWin/HW_329327.pdf).
- [4] E. Hossain, M. Rasti, H. Tabassum, A. Abdelnasser, Evolution toward 5G multi-tier cellular wireless networks: An interference management perspective, *IEEE Wirel. Commun.* 21 (3) (2014) 118–127.
- [5] F.W. Vook, A. Ghosh, T.A. Thomas, MIMO and beamforming solutions for 5G technology, in: 2014 IEEE MTT-S International Microwave Symposium, IMS2014, 2014, pp. 1–4.
- [6] M. Giordani, M. Polese, A. Roy, D. Castor, M. Zorzi, A tutorial on beam management for 3GPP NR at mmWave frequencies, *IEEE Commun. Surv. Tutor.* 21 (1) (2019) 173–196.
- [7] T. Nitsche, C. Cordeiro, A.B. Flores, E.W. Knightly, E. Perahia, J.C. Widmer, IEEE 802.11ad: Directional 60 GHz communication for multi-gigabit-per-second Wi-Fi, *IEEE Commun. Mag.* 52 (12) (2014) 132–141.
- [8] 3GPP, Study on New Radio (NR) Access Technology - Physical Layer Aspects - Release 14, 2017, TR 38.802.
- [9] I.A. Hemadeh, K. Satyanarayana, M. El-Hajjar, L. Hanzo, Millimeter-wave communications: Physical channel models, design considerations, antenna constructions, and link-budget, *IEEE Commun. Surv. Tutor.* 20 (2) (2018) 870–913.
- [10] R. Mailloux, Phased Array Antenna Handbook, third ed., 2017.
- [11] J. Zhang, X. Yu, K.B. Letaief, Hybrid beamforming for 5G and beyond millimeter-wave systems: A holistic view, *IEEE Open J. Commun. Soc.* 1 (2020) 77–91.
- [12] A. Elnashar, S.M. Elnoubi, H.A. El-Mikati, Further study on robust adaptive beamforming with optimum diagonal loading, *IEEE Trans. Antennas and Propagation* 54 (12) (2006) 3647–3658.
- [13] R. Mudumbai, D.R.B. Iii, U. Madhow, H.V. Poor, Distributed transmit beamforming: Challenges and recent progress, *IEEE Commun. Mag.* 47 (2) (2009) 102–110.
- [14] T. Kebede, Y. Wondie, J. Steinbrunn, Channel estimation and beamforming techniques for mm wave-massive MIMO: Recent trends, challenges and open issues, in: 2021 International Symposium on Networks, Computers and Communications, ISNCC, IEEE, 2021, pp. 1–8.
- [15] N. Soltani, K. Sankhe, S. Ioannidis, D. Jaisinghani, K. Chowdhury, Spectrum awareness at the edge: Modulation classification using smartphones, in: 2019 IEEE International Symposium on Dynamic Spectrum Access Networks, DySPAN, 2019, pp. 1–10.
- [16] K. Sankhe, M. Belgiovine, F. Zhou, L. Angioloni, F. Restuccia, S. D'Oro, T. Melodia, S. Ioannidis, K. Chowdhury, No radio left behind: Radio fingerprinting through deep learning of physical-layer hardware impairments, *IEEE Trans. Cogn. Commun. Netw.* 6 (1) (2020) 165–178.
- [17] D. Roy, T. Mukherjee, M. Chatterjee, E. Blasch, E. Pasilio, RFAL: Adversarial learning for RF transmitter identification and classification, *IEEE Trans. Cogn. Commun. Netw.* 6 (2) (2020) 783–801.
- [18] A. Klautau, P. Batista, N. González-Prelcic, Y. Wang, R.W. Heath, 5G MIMO data for machine learning: Application to beam-selection using deep learning, in: 2018 Information Theory and Applications Workshop, ITA, 2018, pp. 1–9.
- [19] Y. Niu, D. Li, L. Jin, A.V. Vasilakos, A survey of millimeter wave communications (mmWave) for 5G: opportunities and challenges, *Wirel. Netw.* 21 (2015) 2657–2676.
- [20] K. Zrar Ghafoor, L. Kong, S. Zeadally, A.S. Sadiq, G. Epiphaniou, M. Ham-moudeh, A.K. Bashir, S. Mumtaz, Millimeter-wave communication for Internet of Vehicles: Status, challenges, and perspectives, *IEEE Internet Things J.* 7 (9) (2020) 8525–8546.
- [21] L. Lu, G.Y. Li, A.L. Swindlehurst, A. Ashikhmin, R. Zhang, An overview of massive MIMO: Benefits and challenges, *IEEE J. Sel. Top. Sign. Proces.* 8 (5) (2014) 742–758.
- [22] K. Zheng, L. Zhao, J. Mei, B. Shao, W. Xiang, L. Hanzo, Survey of large-scale MIMO systems, *IEEE Commun. Surv. Tutor.* 17 (3) (2015) 1738–1760.
- [23] S. Kutty, D. Sen, Beamforming for millimeter wave communications: An inclusive survey, *IEEE Commun. Surv. Tutor.* 18 (2) (2016) 949–973.
- [24] A.F. Molisch, V.V. Ratnam, S. Han, Z. Li, S.L.H. Nguyen, L. Li, K. Haneda, Hybrid beamforming for massive MIMO: A survey, *IEEE Commun. Mag.* 55 (9) (2017) 134–141.
- [25] I. Ahmed, H. Khammari, A. Shahid, A. Musa, K.S. Kim, E. De Poorter, I. Moerman, A survey on hybrid beamforming techniques in 5G: Architecture and system model perspectives, *IEEE Commun. Surv. Tutor.* 20 (4) (2018) 3060–3097.
- [26] pi-radio, Pi-Radio SDR, 2021, <https://www.pi-rad.io/home>.
- [27] NI, What is the mmWave transceiver system (MTS)?, 2021, <https://www.ni.com/en-us/shop/wireless-design-test/what-is-mmwave-transceiver-system.html>.
- [28] Facebook Inc., Terragraph, 2021, <https://terragraph.com/>.
- [29] X. Gu, D. Liu, C. Baks, O. Tageman, B. Sadhu, J. Hallin, L. Rexberg, A. Valdes-Garcia, A multilayer organic package with 64 dual-polarized antennas for 28GHz 5G communication, in: 2017 IEEE MTT-S International Microwave Symposium, IMS, 2017, pp. 1899–1901.
- [30] Y.-N.R. Li, B. Gao, X. Zhang, K. Huang, Beam management in millimeter-wave communications for 5G and beyond, *IEEE Access* 8 (2020) 13282–13293.
- [31] B. Sadhu, Y. Touse, J. Hallin, S. Sahl, S.K. Reynolds, ö. Renström, K. Sjögren, O. Haapalahti, N. Mazar, B. Bokinge, G. Weibull, H. Bengtsson, A. Carlinger, E. Westesson, J. Thillberg, L. Rexberg, M. Yeck, X. Gu, M. Ferriss, D. Liu, D. Friedman, A. Valdes-Garcia, A 28-GHz 32-element TRX phased-array IC with concurrent dual-polarized operation and orthogonal phase and gain control for 5G communications, *IEEE J. Solid-State Circuits* 52 (12) (2017) 3373–3391.

- [32] E. Bjornson, L. Van der Perre, S. Buzzi, E.G. Larsson, Massive MIMO in sub-6 GHz and mmWave: Physical, practical, and use-case differences, *IEEE Wirel. Commun. Mag.* 26 (2) (2019) 100–108.
- [33] R.N. Mitra, D.P. Agrawal, 5G mobile technology: A survey, *ICT Express* 1 (3) (2015) 132–137, Special Issue on Next Generation (5G/6G) Mobile Communications.
- [34] W. Roh, J. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, F. Aryanfar, Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results, *IEEE Commun. Mag.* 52 (2) (2014) 106–113.
- [35] I.A. Remyancev, A.S. Korotkov, Survey on beamforming techniques and integrated circuits for 5G systems, in: 2019 IEEE International Conference on Electrical Engineering and Photonics, EEPolytech, 2019, pp. 76–80.
- [36] H. He, C.K. Wen, S. Jin, G.Y. Li, Deep Learning-Based Channel Estimation for BeamSpace mmWave Massive MIMO Systems, *IEEE Wirel. Commun. Lett.* 7 (5) (2018) 852–855.
- [37] H. Huang, G. Gui, H. Sari, F. Adachi, Deep learning for super-resolution DOA estimation in massive MIMO systems, in: IEEE Vehicular Technology Conference, Vol. 2018-August, no. 9, IEEE, 2018, pp. 8549–8560.
- [38] E. Balevi, A. Doshi, J.G. Andrews, Massive MIMO channel estimation with an untrained deep neural network, *IEEE Trans. Wireless Commun.* 19 (3) (2020) 2079–2090.
- [39] M. Belgiovine, K. Sankhe, C. Bocanegra, D. Roy, K.R. Chowdhury, Deep Learning at the Edge for Channel Estimation in Beyond-5G Massive MIMO, *IEEE Wirel. Commun.* (2021).
- [40] M. Giordani, M. Polese, A. Roy, D. Castor, M. Zorzi, A tutorial on beam management for 3GPP NR at mmWave frequencies, *IEEE Commun. Surv. Tutor.* 21 (1) (2018) 173–196.
- [41] B. Salehi, G. Reus-Muns, D. Roy, Z. Wang, T. Jian, J. Dy, S. Ioannidis, K. Chowdhury, Deep learning on multimodal sensor data at the wireless edge for vehicular network, *IEEE Trans. Veh. Technol.* 71 (7) (2022) 7639–7655.
- [42] P.-H. Kuo, H.T. Kung, P.-A. Ting, Compressive sensing based channel feedback protocols for spatially-correlated massive antenna arrays, in: 2012 IEEE Wireless Communications and Networking Conference, WCNC, 2012, pp. 492–497.
- [43] X. Rao, V.K.N. Lau, Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems, *IEEE Trans. Signal Process.* 62 (12) (2014) 3261–3271.
- [44] C.A. Metzler, A. Maleki, R.G. Baraniuk, From denoising to compressed sensing, *IEEE Trans. Inform. Theory* 62 (9) (2016) 5117–5144.
- [45] C.K. Wen, W.T. Shih, S. Jin, Deep learning for massive MIMO CSI feedback, *IEEE Wirel. Commun. Lett.* 7 (2018) 748–751.
- [46] Y. Yaman, P. Spasojevic, Reducing the LOS Ray Beamforming Setup Time for IEEE 802.11 ad and IEEE 802.15. 3c, in: MILCOM 2016-2016 IEEE Military Communications Conference, IEEE, 2016, pp. 448–453.
- [47] Y. Kishiyama, A. Benjebbour, T. Nakamura, H. Ishii, Future steps of LTE-A: Evolution toward integration of local area and wide area systems, *IEEE Wirel. Commun. Mag.* 20 (1) (2013) 12–18.
- [48] N. González-Prelcic, R. Méndez-Rial, R.W. Heath, Radar Aided Beam Alignment in mmWave V2I Communications Supporting Antenna Diversity, in: 2016 Information Theory and Applications Workshop, ITA, IEEE, 2016, pp. 1–7.
- [49] A. Ali, N. González-Prelcic, A. Ghosh, Passive Radar at the Roadside Unit to Configure Millimeter Wave Vehicle-to-Infrastructure Links, *IEEE Trans. Veh. Technol.* 69 (12) (2020) 14903–14917.
- [50] G. Reus, K. Vijay, C. Bocanegra, Y.C. Eldar, K.R. Chowdhury, Beam Alignment and Tracking for Autonomous Vehicular Communication using IEEE 802.11ad-based Radar, in: IEEE Conference on Computer Communications Workshops, INFOCOM WKSHOPS, IEEE, 2019.
- [51] T. Nitsche, A.B. Flores, E.W. Knightly, J. Widmer, Steering with eyes closed: Mm-Wave beam steering without in-band measurement, in: Proceedings - IEEE INFOCOM, Vol. 26, IEEE, 2015, pp. 2416–2424.
- [52] A. Ali, N. González-Prelcic, R.W. Heath, Millimeter Wave Beam-selection using Out-of-band Spatial Information, *IEEE Trans. Wireless Commun.* 17 (2) (2017) 1038–1052.
- [53] M. Hashemi, C.E. Koksai, N.B. Shroff, Out-of-Band Millimeter Wave Beamforming and Communications to Achieve Low Latency and High Energy Efficiency in 5G Systems, *IEEE Trans. Commun.* 66 (2) (2018) 875–888.
- [54] E.M. Mohamed, H. Kusano, K. Sakaguchi, S. Sampei, WiFi assisted multi-WiGig AP coordination for future multi-Gbps WLANs, in: 2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications, PIMRC, IEEE, 2015.
- [55] A. Ali, N. González-Prelcic, R.W. Heath, Spatial covariance estimation for millimeter wave hybrid systems using out-of-band information, *IEEE Trans. Wireless Commun.* 18 (12) (2019) 5471–5485.
- [56] U. Demirhan, A. Alkhateeb, Radar aided 6G beam prediction: Deep learning algorithms and real-world demonstration, in: 2022 IEEE Wireless Communications and Networking Conference, WCNC, IEEE, 2022, pp. 2655–2660.
- [57] A. Ali, N. González-Prelcic, R.W. Heath, A. Ghosh, Leveraging sensing at the infrastructure for mmWave communication, *IEEE Commun. Mag.* 58 (7) (2020) 84–89.
- [58] J. Choi, V. Va, N. Gonzalez-Prelcic, R. Daniels, C.R. Bhat, R.W. Heath, Millimeter-wave vehicular communication to support massive automotive sensing, *IEEE Commun. Mag.* 54 (12) (2016) 160–167.
- [59] Yole Développement, [Online]. Available: <http://www.yole.fr>.
- [60] Y. Bisheng, L. Fuxun, H. Ronggang, Progress, challenges and perspectives of 3D LiDAR point cloud processing, *Acta Geodaetica Et Cartographica Sinica* 46 (10) (2017) 1509.
- [61] G. Reus-Muns, B. Salehi, D. Roy, T. Jian, Z. Wang, J. Dy, S. Ioannidis, K. Chowdhury, Deep learning on visual and location data for V2I mmWave beamforming, in: 2021 17th International Conference on Mobility, Sensing and Networking, MSN, IEEE, 2021, pp. 559–566.
- [62] A. Capone, I. Filippini, V. Sciancalepore, D. Tremolada, Obstacle avoidance cell discovery using mm-waves directive antennas in 5G networks, in: 2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications, PIMRC, IEEE, 2015, pp. 2349–2353.
- [63] J. Kim, A.F. Molisch, Enabling gigabit services for IEEE 802.11 ad-capable high-speed train networks, in: 2013 IEEE Radio and Wireless Symposium, IEEE, 2013, pp. 145–147.
- [64] V. Va, T. Shimizu, G. Bansal, R.W. Heath, Beam design for beam switching based millimeter wave vehicle-to-infrastructure communications, in: 2016 IEEE International Conference on Communications, ICC, IEEE, 2016, pp. 1–6.
- [65] Y. Wang, A. Klautau, M. Ribero, M. Narasimha, R.W. Heath, MmWave Vehicular Beam Training with Situational Awareness by Machine Learning, in: 2018 IEEE Globecom Workshops, GC Wkshps, 2018, pp. 1–6.
- [66] V. Va, J. Choi, T. Shimizu, G. Bansal, R.W. Heath, Inverse Multipath Fingerprinting for Millimeter Wave V2I Beam Alignment, *IEEE Trans. Veh. Technol.* 67 (5) (2017) 4042–4058.
- [67] G.H. Sim, S. Klos, A. Asadi, A. Klein, M. Hollick, An online context-aware machine learning algorithm for 5G mmWave vehicular communications, *IEEE/ACM Trans. Netw.* 26 (6) (2018) 2487–2500.
- [68] J.C. Aviles, A. Kouki, Position-aided mm-wave beam training under NLOS conditions, *IEEE Access* 4 (2016) 8703–8714.
- [69] M. Alrabeiah, J. Booth, A. Hredzak, A. Alkhateeb, ViWi Vision-Aided mmWave Beam Tracking: Dataset, Task, and Baseline Solutions, 2020, arXiv preprint arXiv:2002.02445.
- [70] R. Dey, F.M. Salem, Gate-variants of gated recurrent unit (GRU) neural networks, in: 2017 IEEE 60th International Midwest Symposium on Circuits and Systems, MWSCAS, IEEE, 2017, pp. 1597–1600.
- [71] D. Marasinghe, N. Jayaweera, N. Rajatheva, S. Hakola, T. Koskela, O. Tervo, J. Karjalainen, E. Tirola, J. Hukkonen, LiDAR aided wireless networks-beam prediction for 5G, in: 2022 IEEE 96th Vehicular Technology Conference, VTC2022-Fall, IEEE, 2022, pp. 1–7.
- [72] Y. Tian, G. Pan, M.-S. Alouini, Applying deep-learning-based computer vision to wireless communications: Methodologies, opportunities, and challenges, *IEEE Open J. Commun. Soc.* (2020).
- [73] S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Comput.* 9 (8) (1997) 1735–1780.
- [74] W. Xu, F. Gao, S. Jin, A. Alkhateeb, 3D scene-based beam selection for mmWave communications, *IEEE Wirel. Commun. Lett.* 9 (11) (2020) 1850–1854.
- [75] B. Salehi, M. Belgiovine, S.G. Sanchez, J. Dy, S. Ioannidis, K. Chowdhury, Machine learning on camera images for fast mmWave beamforming, in: 2020 IEEE 17th International Conference on Mobile Ad Hoc and Sensor Systems, MASS, IEEE, 2020, pp. 338–346.
- [76] M.K. Haider, Y. Ghasempour, D. Koutsonikolas, E.W. Knightly, LiSteer: MmWave beam acquisition and steering by tracking indicator LEDs on wireless APs, in: Proceedings of the 24th Annual International Conference on Mobile Computing and Networking, 2018, pp. 273–288.
- [77] T. Woodford, X. Zhang, E. Chai, K. Sundaresan, A. Khojastepour, SpaceBeam: LiDAR-driven one-shot mmWave beam management, in: Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services, 2021, pp. 389–401.
- [78] Microsoft, 2020, Azure Kinect., <https://azure.microsoft.com>.
- [79] Airfide, 2020, <https://airfidenet.com/>.
- [80] S. Jiang, G. Charan, A. Alkhateeb, LiDAR Aided Future Beam Prediction in Real-World Millimeter Wave V2I Communications, *IEEE Wirel. Commun. Lett.* (2022).
- [81] A. Alkhateeb, G. Charan, T. Osman, A. Hredzak, J. Morais, U. Demirhan, N. Srinivas, DeepSense 6G: A large-scale real-world multi-modal sensing and communication dataset, 2022, arXiv preprint arXiv:2211.09769.
- [82] M. Alrabeiah, A. Hredzak, Z. Liu, A. Alkhateeb, ViWi: A deep learning dataset framework for vision-aided wireless communications, in: 2020 IEEE 91st Vehicular Technology Conference, VTC2020-Spring, IEEE, 2020, pp. 1–5.
- [83] Remcom, “Wireless Insite,” <http://www.remcom.com/wireless-insite>.
- [84] Blender, <https://www.blender.org>.
- [85] R. Heinzler, P. Schindler, J. Seekircher, W. Ritter, W. Stork, Weather Influence and Classification with Automotive LiDAR Sensors, in: 2019 IEEE Intelligent Vehicles Symposium, IV, 2019, pp. 1527–1534.
- [86] A. Klautau, N. González-Prelcic, R.W. Heath, LiDAR Data for Deep Learning-Based mmWave Beam-Selection, *IEEE Wirel. Commun. Lett.* 8 (3) (2019) 909–912.

- [87] M. Dias, A. Klautau, N. González-Prelcic, R.W. Heath, Position and LIDAR-aided mmWave beam selection using deep learning, in: 2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications, SPAWC, IEEE, 2019, pp. 1–5.
- [88] M. Alrabeiah, A. Hredzak, A. Alkhateeb, Millimeter Wave Base Stations with Cameras: Vision-Aided Beam and Blockage Prediction, in: 2020 IEEE 91st Vehicular Technology Conference, VTC2020, IEEE, 2020, pp. 1–5.
- [89] G. Charan, A. Hredzak, C. Stoddard, B. Berrey, M. Seth, H. Nunez, A. Alkhateeb, Towards real-world 6G drone communication: Position and camera aided beam prediction, in: 2022 IEEE Global Communications Conference, GLOBECOM, 2022.
- [90] G. Charan, T. Osman, A. Hredzak, N. Thawdar, A. Alkhateeb, Vision-position multi-modal beam prediction using real millimeter wave datasets, in: 2022 IEEE Wireless Communications and Networking Conference, WCNC, 2022.
- [91] B. Salehi, J. Gu, D. Roy, K. Chowdhury, FLASH: Federated learning for automated selection of high-band mmWave sectors, in: IEEE INFOCOM 2022-IEEE Conference on Computer Communications, IEEE, 2022, pp. 1719–1728.
- [92] J. Gu, B. Salehi, D. Roy, K.R. Chowdhury, Multimodality in mmWave MIMO beam selection using deep learning: Datasets and challenges, IEEE Commun. Mag. 60 (11) (2022) 36–41.
- [93] D. Steinmetzer, D. Wegemer, M. Schulz, J. Widmer, M. Hollick, Compressive Millimeter-Wave Sector Selection in Off-the-Shelf IEEE 802.11ad Devices, in: International Conference on Emerging Networking EXperiments and Technologies, CoNEXT, 2017.
- [94] FLASH Dataset, <https://genesys-lab.org/multimodal-fusion-nextg-v2x-communications>.
- [95] P.A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, E. Wiessner, Microscopic Traffic Simulation using SUMO, in: International Conference on Intelligent Transportation Systems, ITSC, 2018, pp. 2575–2582.
- [96] M. Gschwandtner, R. Kwitt, A. Uhl, W. Pree, BlenSor: Blender sensor simulation toolbox, in: Advances in Visual Computing, Springer Berlin Heidelberg, 2011, pp. 199–208.
- [97] MarketsandMarkets Research Private Ltd, V2X Market for Automotive by Communication Type, 2020, <https://www.marketsandmarkets.com/Market-Reports/automotive-vehicle-toeverything-v2x-market-90013236.html>.
- [98] CFI Education Inc, Vehicle to everything (V2X), 2021, <https://www.corporatefinanceinstitute.com/resources/knowledge/other/vehicle-to-everything-v2x/>.
- [99] I. Maskulainen, P. Luoto, P. Pirinen, M. Bennis, K. Horneman, M. Latva-aho, Performance evaluation of adaptive beamforming in 5G-V2X networks, in: European Conference on Networks and Communications, EuCNC, 2017, pp. 1–5.
- [100] G.H. Lee, J.D. Choi, J.H. Lee, M.Y. Kim, Object detection using vision and LiDAR sensor fusion for multi-channel V2X system, in: 2020 International Conference on Artificial Intelligence in Information and Communication, ICAIC, 2020, pp. 1–5.
- [101] M.A. Khan, A. Safi, I.M. Qureshi, I.U. Khan, Flying ad-hoc networks (FANETs): A review of communication architectures, and routing protocols, in: 2017 First International Conference on Latest Trends in Electrical Engineering and Computing Technologies, INTELECT, 2017, pp. 1–9.
- [102] Y. Tian, G. Pan, M.-S. Alouini, Applying Deep-Learning-Based Computer Vision to Wireless Communications: Methodologies, Opportunities, and Challenges, IEEE Open J. Commun. Soc. 2 (2021).
- [103] K. Alemdar, D. Varshey, S. Mohanti, U. Muncuk, K.R. Chowdhury, RFClock: Timing, phase and frequency synchronization for distributed wireless networks, in: MobiCom, 2021, pp. 15–27.
- [104] S. Mohanti, C. Bocanegra, J. Meyer, G. Secinti, M. Diddi, H. Singh, K. Chowdhury, AirBeam: Experimental demonstration of distributed beamforming by a swarm of UAVs, in: 2019 IEEE 16th International Conference on Mobile Ad Hoc and Sensor Systems, MASS, IEEE, 2019, pp. 162–170.
- [105] S.G. Sanchez, K.R. Chowdhury, Robust 60ghz beamforming for UAVs: Experimental analysis of hovering, blockage and beam selection, IEEE Internet Things J. (2020) 1.
- [106] S. Garcia Sanchez, S. Mohanti, D. Jaisinghani, K.R. Chowdhury, Millimeter-wave base stations in the sky: An experimental study of UAV-to-ground communications, IEEE Trans. Mob. Comput. (2020) 1.
- [107] J.A. Marshall, A. Bonchis, E. Nebot, S. Scheduling, Robotics in mining, in: Springer Handbook of Robotics, Springer, 2016, pp. 1549–1576.
- [108] S. Bonadies, A. Lefcourt, S.A. Gadsden, A survey of unmanned ground vehicles with applications to agricultural and environmental sensing, in: Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping, Vol. 9866, International Society for Optics and Photonics, 2016, p. 98660Q.
- [109] S. Young, A. Kott, A survey of research on control of teams of small robots in military operations, 2016, arXiv preprint arXiv:1606.01288.
- [110] L. Pedersen, D. Kortenkamp, D. Wettergreen, I. Nourbakhsh, A survey of space robotics, in: Proceedings of the 7th International Symposium on Artificial Intelligence, Robotics and Automation in Space, 2003, pp. 19–23.
- [111] J. Liu, J. Wu, Multiagent Robotic Systems, CRC Press, 2018.
- [112] K. Brink, R. Sherrill, J. Godwin, J. Zhang, A. Willis, Maplets: An efficient approach for cooperative SLAM map building under communication and computation constraints, in: 2020 IEEE/ION Position, Location and Navigation Symposium, PLANS, IEEE, 2020, pp. 367–374.
- [113] M.U.M. Bhutta, M. Kuse, R. Fan, Y. Liu, M. Liu, Loop-box: Multiagent direct slam triggered by single loop closure for large-scale mapping, IEEE Trans. Cybern. (2020).
- [114] K. Krinkin, A. Filatov, A. Filatov, Modern multi-agent slam approaches survey, in: Proceedings of the XXth Conference of Open Innovations Association FRUCT, Vol. 776, 2017, pp. 617–623.
- [115] J.J. Chung, C. Rebhuhn, C. Yates, G.A. Hollinger, K. Tumer, A multiagent framework for learning dynamic traffic management strategies, Auton. Robots 43 (6) (2019) 1375–1391.
- [116] D.B. Poudel, Coordinating hundreds of cooperative, autonomous robots in a warehouse, Jan 27 (1–13) (2013) 26.
- [117] M. Fujita, Y. Domae, A. Noda, G. Garcia Ricardez, T. Nagatani, A. Zeng, S. Song, A. Rodriguez, A. Causo, I.-M. Chen, et al., What are the important technologies for bin picking? Technology analysis of robots in competitions based on a set of performance metrics, Adv. Robot. 34 (7–8) (2020) 560–574.
- [118] J. Bodner, H. Wykypiel, G. Wetscher, T. Schmid, First experiences with the da Vinci™ operating robot in thoracic surgery, Eur. J. Cardio-Thoracic Surg. 25 (5) (2004) 844–851.
- [119] P.P. Rao, Robotic surgery: New robots and finally some real competition! World J. Urol. 36 (4) (2018) 537–541.
- [120] S. Shalev-Shwartz, S. Shammah, A. Shashua, Safe, multi-agent, reinforcement learning for autonomous driving, 2016, arXiv preprint arXiv:1610.03295.
- [121] C. Yu, X. Wang, X. Xu, M. Zhang, H. Ge, J. Ren, L. Sun, B. Chen, G. Tan, Distributed multiagent coordinated learning for autonomous driving in highways based on dynamic coordination graphs, IEEE Trans. Intell. Transp. Syst. 21 (2) (2019) 735–748.
- [122] P. Palanisamy, Multi-agent connected autonomous driving using deep reinforcement learning, in: 2020 International Joint Conference on Neural Networks, IJCNN, IEEE, 2020, pp. 1–7.
- [123] V. Arguenon, A. Bergues-Lagarde, C. Rosenberger, P. Bro, W. Smari, Multi-agent based prototyping of agriculture robots, in: International Symposium on Collaborative Technologies and Systems, CTS'06, IEEE, 2006, pp. 282–288.
- [124] M. Afrin, J. Jin, A. Rahman, A. Rahman, J. Wan, E. Hossain, Resource allocation and service provisioning in multi-agent cloud robotics: A comprehensive survey, IEEE Commun. Surv. Tutor. (2021).
- [125] M. Davoodi, S. Faryadi, J.M. Velni, A graph theoretic-based approach for deploying heterogeneous multi-agent systems with application in precision agriculture, J. Intell. Robot. Syst. 101 (1) (2021) 1–15.
- [126] C.X. Lu, S. Rosa, P. Zhao, B. Wang, C. Chen, J.A. Stankovic, N. Trigoni, A. Markham, See through smoke: Robust indoor mapping with low-cost mmWave radar, in: Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services, 2020, pp. 14–27.
- [127] H. Elayan, O. Amin, R.M. Shubair, M.-S. Alouini, Terahertz communication: The opportunities of wireless technology beyond 5G, in: 2018 International Conference on Advanced Communication Technologies and Networking, CommNet, 2018, pp. 1–5.
- [128] H.-J. Song, T. Nagatsuma, Present and future of terahertz communications, IEEE Trans. Terahertz Sci. Technol. 1 (1) (2011) 256–263.
- [129] C. Lin, G.Y. Li, L. Wang, Subarray-based coordinated beamforming training for mmWave and Sub-THz communications, IEEE J. Sel. Areas Commun. 35 (9) (2017) 2115–2126.



**Debashri Roy** received her MS (2018) and Ph.D. (2020) degrees in Computer Science from University of Central Florida, USA. She is currently an associate research scientist at Northeastern University. Her research interests are in the areas of AI/ML enabled technologies in wireless communication, multimodal data fusion, network orchestration and nextG networks.



**Batool Salehi** is currently pursuing a Ph.D. degree in computer engineering at Northeastern University under the supervision of Prof. K. Chowdhury. She received her M.S. (2019) in Electrical Engineering from University of Tehran, Iran. Her current research focuses on mmWave beamforming, Internet of Things, and the application of machine learning in the domain of wireless communication.





**Stella Banou** received her B.Sc. in Electrical and Computer Engineering (ECE) at Worcester Polytechnic Institute in Worcester, MA in 2016. She joined Northeastern University's GENESYS lab in 2016 as part of the M.Sc./PhD program and is currently a Ph.D. candidate working with Prof. Kaushik Chowdhury. Her research is on alternative intra-body communication methods (Galvanic Coupling, Magnetic Resonance Coupling), wireless charging for implants, intra-body sensor networks and applications of healthcare IoT.



**Prashant Ganesh** is currently a Research Assistant Engineer at the University of Florida. He obtained his Bachelors' degree in Electrical and Electronics Engineering from Anna University, India and his Masters' degrees in Aerospace engineering from the University of Colorado at Boulder. His research interests is in developing robust estimation and control techniques for autonomous agents operating in an GPS denied region.



**Subhramoy Mohanti** is a Ph.D. candidate at the Department of Electrical and Computer Engineering in Northeastern University. He is the recipient of the IEEE INFOCOM Best Paper Award (2018), the Northeastern University Graduate Dissertation Research Grant (2015) for his Master's thesis and NSF Travel Grant for IEEE INFOCOM 2018. His current research areas include aerial distributed beamforming, machine learning enabled UAV identification, UAV networks, designing next generation wireless protocols and networks, along with scheduling and optimization. He has published and co-authored many papers on these topics. He has also served as reviewer for prestigious journals and conferences on these topics.



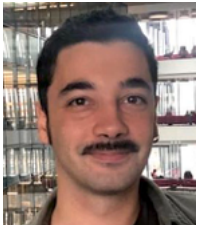
**Chris Dickis** is a wireless architect at NVIDIA and the technical lead for the application of AI and machine learning to 5G and 6G wireless. In his 24 years working in signal processing and communications he has delivered silicon and software products for 3G, 4G, and 5G baseband DSP and Docsis 3.1 cable access. He has performed research and delivered products for digital frontend (DFE) technology for cellular systems with a particular emphasis on digital predistortion for power amplifier linearization. Chris has also worked extensively on architectures and compilers for machine learning in computer vision. Before moving to Silicon Valley in 1998, he was a tenured academic in Melbourne Australia for 13 years. He has over 150 publications and 70 patents and is an adjunct professor at Santa Clara University, where he has taught courses on real-time signal processing and machine learning for 18 years. In 2018, he was awarded the IEEE Communications Society Award for Advances in Communication for research in full-duplex wireless communication.



**Guillem Reus-Muns** received a B.Sc. degree in telecommunications engineering from the Polytechnic University of Catalonia (UPC-BarcelonaTech). He joined Northeastern University, USA, in 2018, where he got his M.Sc. in electrical and computer engineering and is currently working towards his Ph.D. His current research interests involve cellular networks, machine learning for wireless communications, and spectrum access.



**Kaushik Roy Chowdhury** (M'09-SM'15) is a Professor at Northeastern University, Boston, MA. His current research interests involve systems aspects of networked robotics, machine learning for agile spectrum sensing/access, wireless energy transfer, and large-scale experimental deployment of emerging wireless technologies. He received the M.S. degree from the University of Cincinnati in 2006, and the Ph.D. degree from the Georgia Institute of Technology in 2009. He is currently a Professor with the Electrical and Computer Engineering Department. He was a winner of the Presidential Early Career Award for Scientists and Engineers (PECASE) in 2017, ONR Director of Research Early Career Award in 2016 and the NSF CAREER Award in 2015. His current research interests involve systems aspects of networked robotics, machine learning for agile spectrum sensing/access, wireless energy transfer, and large-scale experimental deployment of emerging wireless technologies.



**Mauro Belgiovine** is pursuing his Ph.D. at the Electrical and Computer Engineering department at Northeastern University, under the guidance of Prof. Kaushik Chowdhury. His current research interests involve deep learning, wireless communication, and heterogeneous computing.