An In-Depth Study of Uplink Performance of 5G mmWave Networks

Moinak Ghoshal[1*], Z. Jonny Kong[2*], Qiang Xu[2*], Zixiao Lu[1], Shivang Aggarwal[1], Imran Khan[1], Yuanjie Li[3], Y. Charlie Hu[2], Dimitrios Koutsonikolas[1]

[1]Institute for the Wireless Internet of Things, Northeastern University, USA,
[2]Purdue University, USA, [3]Tsinghua University, China

ABSTRACT

The highly anticipated 5G mmWave technology promises to enable many uplink-oriented, latency-critical applications (LCAs) such as Augmented Reality and Connected Autonomous Vehicles. Nonetheless, recent measurement studies have largely focused on its downlink performance. In this work, we perform a systematic study of the *uplink* performance of commercial 5G mmWave networks across 3 major US cities and 2 mobile operators. Our study makes three contributions. (1) It reveals that 5G mmWave uplink performance is geographically diverse, substantially higher over LTE in terms of bandwidth and latency, but often erratic and suboptimal, which can degrade LCA performance. (2) Our analysis of control messages and PHY-level KPIs shows that the root causes for the suboptimal performance are fundamental to 5G mmWave and cannot be easily fixed via simple tuning of network configurations. (3) We identify various design and deployment optimizations that 5G operators can explore to bring 5G mmWave performance to the level needed to ultimately support the LCAs.

CCS CONCEPTS

• **Networks → Mobile networks**; **Network measurement**; **Network performance analysis**.

ACM Reference Format:
Moinak Ghoshal, Z. Jonny Kong, Qiang Xu, Shivang Aggarwal, Imran Khan, Yuanjie Li, Y. Charlie Hu, Dimitrios Koutsonikolas. 2022. An In-Depth Study of Uplink Performance of 5G mmWave Networks. In *ACM SIGCOMM 2022 Workshop on 5G and Beyond Network Measurements, Modeling, and Use Cases (5G-MeMU '22), August 22, 2022, Amsterdam, Netherlands.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3538394.3546042

1 INTRODUCTION

The evolution of mobile networks has come a long way. The most recent generation of cellular networks, 5G, and in particular, 5G mmWave, promises unprecedented high bandwidth and ultra low latency and holds the promise to finally support latency-critical applications (LCAs) such as Augment Reality (AR), Mixed Reality (XR), and Connected Autonomous Vehicles (CAVs) [30]. Such applications are *latency-critical* because *their core tasks need to be completed in real time* in order to satisfy their stringent Quality-of-Experience (QoE) requirements. For example, AR applications need to perform computationally heavy vision tasks, such as object detection, pose estimation, and depth estimation, at low latency and high frame rate (e.g., 30 or 60 FPS). As mobile devices have constrained compute capabilities, they are forced to offload such tasks to powerful edge cloud over the wireless network in real time. Similarly, for CAVs, the individual vehicles have to exchange real-time, high-fidelity sensor data over the wireless network. Such stringent QoE requirements of LCAs in turn place high *uplink* bandwidth demand on the wireless access network.

While 5G deployments are still in the early stage, a number of measurement studies have been conducted to assess their performance and impact on applications [26–29]. These studies found that, although today's mmWave deployments indeed offer Gbps throughput and lower latency than 4G LTE, their performance is often suboptimal, coverage is sporadic, the handover process is not optimized, and applications cannot always take advantage of the full potential of 5G mmWave. However, these studies have mostly focused on measuring the 5G *downlink* performance and the uplink performance of 5G networks remains largely unknown.

In this work, we aim to fill this gap by answering the question: *How good is 5G network uplink performance?* More specifically, since most LCAs distinguish themselves from legacy apps for their heavy, bursty *uplink* data transfer, can current 5G deployments meet the uplink traffic demand of LCAs, as 5G has provisioned much higher downlink than uplink bandwidth similarly as all its predecessors?

To answer this question, in this work, we conduct to our knowledge, the first detailed study of uplink performance over commercial 5G mmWave networks. Through an extensive measurement campaign across 3 US cities (Boston, Chicago, and Indianapolis) and 2 mobile operators (Verizon and AT&T), we shed light on how diverse the 5G mmWave uplink performance can be in different cities. Our study covers diverse scenarios like static, walking, and driving, and compares the 5G mmWave results with those for the 4G LTE network. Our results show that 5G mmWave can achieve 3x gain over LTE in terms of uplink bandwidth and latency for the baseline scenario, where the user equipment (UE) is facing towards the base station (BS), but the performance deteriorates to be almost comparable to or worse than that for 4G LTE for more challenging scenarios, where the UE faces away from the BS or when the user is walking or driving. Our close examination of the results at the millisecond time scale shows that 5G mmWave uplink suffers serious degradation and high throughput fluctuations due to blockage and UE mobility.

[*]Moinak Ghoshal, Z. Jonny Kong, and Qiang Xu are co-primary authors.

5G-MeMU '22, August 22, 2022, Amsterdam, Netherlands

ACM ISBN 978-1-4503-9393-5/22/08…$15.00
https://doi.org/10.1145/3538394.3546042

(a) Boston.

(b) Chicago.

(c) Indianapolis.

Figure 1: 5G mmWave deployment in different cities.

We further uncover the root causes for the throughput degradation by analyzing network control messages and PHY-level KPIs, such as MCS and stability of beamforming, by extending MobileInsight [20, 21] to support various 5G mmWave-specific messages (we open-sourced these enhancements to MobileInsight). Our analysis shows that the root causes for the low and highly fluctuating 5G uplink bandwidth are fundamental to 5G mmWave and cannot be easily fixed via simple tuning of network configurations.

Finally, we discuss how network designers and operators can innovate on 5G mmWave designs (which call for architecture/protocol changes), such as device-centric uplink mmWave, QoE-driven mmWave adaptation, and blockage and mobility mitigation by prediction, and on operations (which can be readily implemented today) on the UE and infrastructure, such as uplink carrier-aggregation, application-aware 5G dual connectivity, and dense cell deployments, to overcome these limitations.

We are making all our measurement data publicly available [**?**] to enable the community to study the feasibility of today's 5G mmWave uplink in supporting LCAs such as AR, real-time video analytics [14], and CAVs.

2 MAIN MEASUREMENT CAMPAIGN

In this section, we describe the results of our main measurement campaign in Boston.

2.1 Methodology

5G Carriers. We used Verizon for our main measurement campaign, since it is the only operator that provides 5G mmWave service in Boston. Verizon's 5G mmWave service works in both 28 and 39 GHz frequency bands (n261/260). We noticed that the Verizon mmWave base stations (BSes) are mounted on the walls of high buildings (see Fig. 1a) in a sparse manner with two adjacent BSes being up to a few miles apart from each other.

5G UE and Cloud Servers. We used a rooted Google Pixel 5 phone as the UE. The Pixel 5 supports the 5G mmWave bands n260/261 (39/28 GHz) and 4-CC (4x100 MHz)/1-CC downlink/uplink carrier aggregation. We used two servers from AWS: a traditional cloud

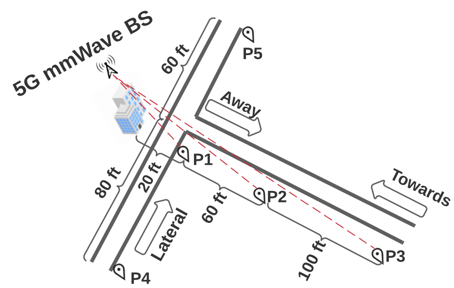


Figure 2: Positions and trajectories for static and walking experiments in Boston.

server and an edge server (AWS Wavelength) [2], to perform the throughput and latency measurements. Wavelength servers are located inside Verizon's network and specially designed for edge computing to provide shorter latency compared to traditional cloud servers. We performed some preliminary upload tests to observe the AWS edge and traditional cloud server performance and found that in terms of throughput, both servers achieve similar values. However, the latency for the edge server was almost half of that of the cloud server.

Hence, in the interest of space, we only show results with the AWS Wavelength edge server for Boston.

Experiments. We used nuttcp [3] with the default TCP congestion control, CUBIC, to generate backlogged TCP uplink traffic to measure the uplink throughput, and we used the ICMP-based ping utility to measure the latency at an interval of 100 ms. To perform a holistic study of the behavior of 5G mmWave uplink, we considered three scenarios in our measurements:

- *Static*: We conducted the static measurements near a university campus at times with minimal human and vehicle presence. This ensured that factors like blockage and background data usage did not affect our measurements. We held the phone at P1, P2, P3 (Fig. 2), while facing towards and away from the BS, and measured the uplink throughput and latency for 15 s.

Table 1: 5G mmWave and LTE performance (mean ± standard deviation) in Boston for various scenarios.

Scenario	Throughput (Mbps)	RTT (ms)
5G Static, P1, toward	131.8±22.3	14.7±1.1
5G Static, P2, toward	153.0±17.3	14.5±0.7
5G Static, P3, toward	157.7±14.7	14.5±0.6
5G Static, P1, away	72.4±30.1	14.9±1.1
5G Static, P2, away	52.8±21.1	15.1±1.3
5G Static, P3, away	83.6±22.0	14.8±1.0
5G Walking, toward	127.9±21.0	14.9±1.1
5G Walking, away	50.5±26.5	15.1±0.9
5G Walking, lateral	59.5±30.5	19.0±8.8
5G Driving	39.7±21.2	33.0±22.0
LTE, Static	53.4±3.2	47.6±8.4
LTE, Walking	53.5±2.8	44.1±7.5
LTE, Driving	33.1±6.2	34.9±17.0

- *Walking*: We walked towards, away from, and laterally to the BS in the same location where we conducted the static measurements (Fig. 2), and took measurements along the way for 40-60 s. In the case of walking towards (away from) the BS, we started at P1 (P3) and stopped at P3 (P1). For lateral motion, we started at P4, walked for 80 ft until reaching the BS, and continued walking for another 40 ft until hitting P5. In all cases, we walked at the typical walking speed (~3 ft/s).
- *Driving*: We collected uplink throughput traces while driving on a moderately congested road in downtown Boston. The trajectory is around 2 miles in length and we drove at a speed of 20 miles per hour, taking 7-8 minutes to complete it. We did the experiments at night after 9 PM to avoid heavy road traffic. However, sometimes, we had to stop at traffic signals for a brief period of time.

We collected 6 traces for each position-direction combination in the static and walking scenarios, and 8 traces for the driving scenario. To mitigate the impact of temporal network conditions, we interleaved measurements for different conditions and spread them across days. In all three scenarios, we also performed the same measurements over LTE using a separate Pixel 5 phone. Table 1 summarizes the results and Figs. 3, 4, 5, 6 show representative throughput timelines.

2.2 Performance under Static Scenario

Facing the BS. This is the best-case scenario for 5G mmWave. We make the following observations. (1) *The average 5G mmWave uplink throughput is about 3x higher than the LTE throughput* (Fig. 3) at all three positions.[1] For comparison, we note that the downlink 5G mmWave throughput is more than 10x higher (1.7 - 1.9 Gbps) at the same locations. In contrast to WiFi mmWave technologies [17], 5G mmWave is highly *asymmetric*. (2) *5G mmWave offers much lower and much more stable latency compared to LTE* (Table 1).

Facing away from the BS. In this scenario, the mmWave signal is partially blocked by the user holding the device. We make the following observations: (1) We notice a *drastic drop in 5G throughput* compared to the scenario where the UE faces the BS (Table 1). (2) In addition to the lower average values, *5G throughput exhibits large fluctuations over time* (Fig. 4). (3) The latency for both technologies remains similar as in the case of facing towards the BS (Table 1).

[1]We conjecture that the lower throughput at P1, the position closest to the BS, is due to poor beamforming capabilities of the UE on the vertical plane.

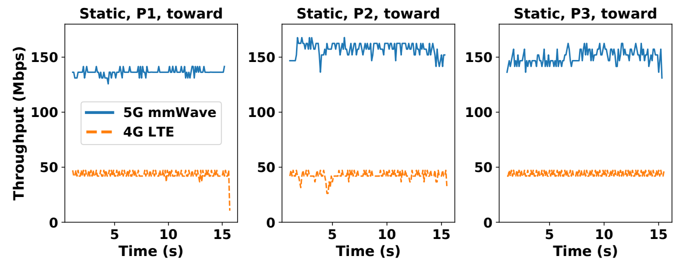


Figure 3: 5G mmWave and LTE throughput at 3 positions while facing the BS.

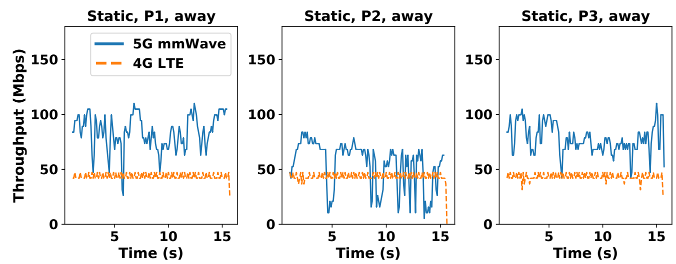


Figure 4: 5G mmWave and LTE throughput at 3 positions while facing away from the BS.

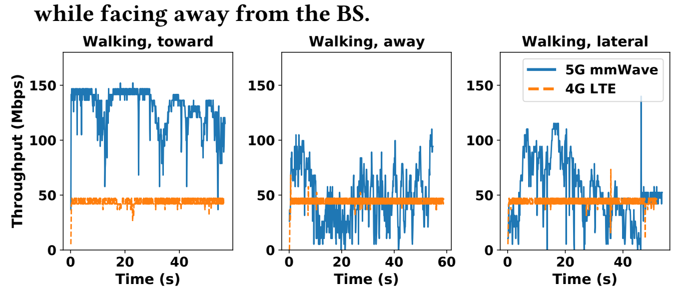


Figure 5: 5G mmWave and LTE throughput – walking.

2.3 Performance under Walking Scenario

We study the impact of mobility alone (moving towards the BS) or in combination with blockage (moving away from the BS and in the case of lateral motion after the user crosses the point in front of the BS). We make the following observations: (1) In general, the average performance of 5G mmWave degrades for all three mobility patterns, compared to the static case when the user faces the BS (Table 1). The drop is more pronounced when the user moves away from or laterally to the BS. (2) The 5G mmWave throughput exhibits even higher fluctuations compared to the static scenario with the user facing away from the BS, *e.g.,* in the case of lateral motion the throughput ranges from 115 Mbps to as low as 0 Mbps (Fig. 5). In contrast, the LTE throughput is not affected by mobility and retains a similar value as in the static scenarios. (3) The latency with both technologies is similar to that of the static scenarios (Table 1), with a few exceptions in the lateral motion for 5G mmWave, where we observed a few spikes up to 150 ms.

2.4 Performance under Driving Scenario

We make the following observations: (1) In general, the average uplink performance degrades and exhibits higher fluctuations for

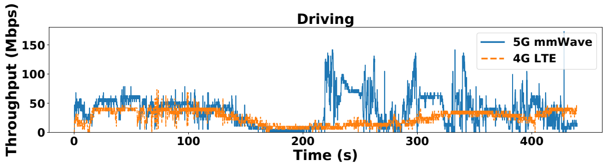


Figure 6: 5G mmWave and LTE throughput – driving.

both 5G mmWave and LTE compared to static and walking scenarios (Table 1 and Fig. 6). We observe noticeable drops in throughput (often down to less than 10 Mbps) for prolonged periods of time, even for LTE (Fig. 6). (2) The average latency for 5G mmWave and LTE are similar while driving (Table 1) with very large fluctuations.

3 DIAGNOSING 5G MMWAVE PERFORMANCE

To diagnose the performance problems of 5G mmWave uplink, we extended MobileInsight [20] (MI) to collect low-level over-the-air 5G mmWave messages between the UE and the BS, including PHY-layer channel measurements, scheduling information, mmWave beam measurements, MAC-layer transport blocks, buffer status reports, RLC layer segment statistics, and RRC signaling messages. In the following, we analyze the impact of channel fluctuations due to blockage and mobility, and the positive/negative impact of beamforming.

Channel fluctuations. We extracted the uplink modulation and coding scheme (MCS) [7] from MI, logged at a granularity of 3 ms. The uplink MCS is calculated by the BS based on runtime channel quality, and sent to devices through the PHY-layer signaling channel. Fig. 7 plots the throughput and MCS timelines for different scenarios. The throughput and MCS have a strong correlation across all scenarios. The throughput fluctuations also closely follow the MCS fluctuations. When the user faces away from the BS or is walking/driving, the 5G MCS drops to lower indices, and exhibits high fluctuations. Hence, we can confirm that MCS (hence channel conditions) is responsible for the poor throughput under 5G mmWave in blockage and mobility scenarios. In contrast, the LTE MCS remains stable across all scenarios except during driving.

Beamforming. To observe the impact of beamforming on 5G mmWave performance, we extracted the beam Synchronization Signal Block (SSB) index [7], which uniquely identifies each beam, from MI, logged at a granularity of 20 ms, and analyzed its temporal dynamics (Fig. 7). The SSB index remains unchanged in static conditions regardless of the user orientation, which implies that beamforming is not triggered. In contrast, the SSB index under mobility varies over time, especially in the walking laterally and driving cases, indicating that the UE and BS try to re-align their beams continuously.

To understand the impact of beam re-alignment on throughput, we calculated the throughput difference before and after (15 ms) each beamforming event and plot the CDF of the differences (Fig. 7i). We observe that around 60% of the time, beamforming has a positive impact on throughput. However, the throughput improvement is mostly small, possibly due to the overhead of discovering a new beam. Improvements of more than 25 Mbps are only observed in 5% of the cases. More importantly, the throughput drops in 25–30% of the cases, due to three possible reasons: (i) the new beam establishes communication over a path with higher path loss when the LOS is blocked, (ii) beamforming is triggered unnecessarily, and (iii) a suboptimal beam is selected as the signal strength constantly changes due to mobility during the beamforming process. For example, in the "walking, lateral" scenario (Fig. 7e), the SSB index oscillates between 43 and 33 in the first 3 s. In the same trace, beam 43 is reused at 38–40 s. This is unlikely to be optimal, as the UE is between the BS and P5 at that time (cf. Fig. 2) and has a very different orientation w.r.t. the BS compared to the first 3 s, when it is near P4. The throughput drops after beamforming in both cases.

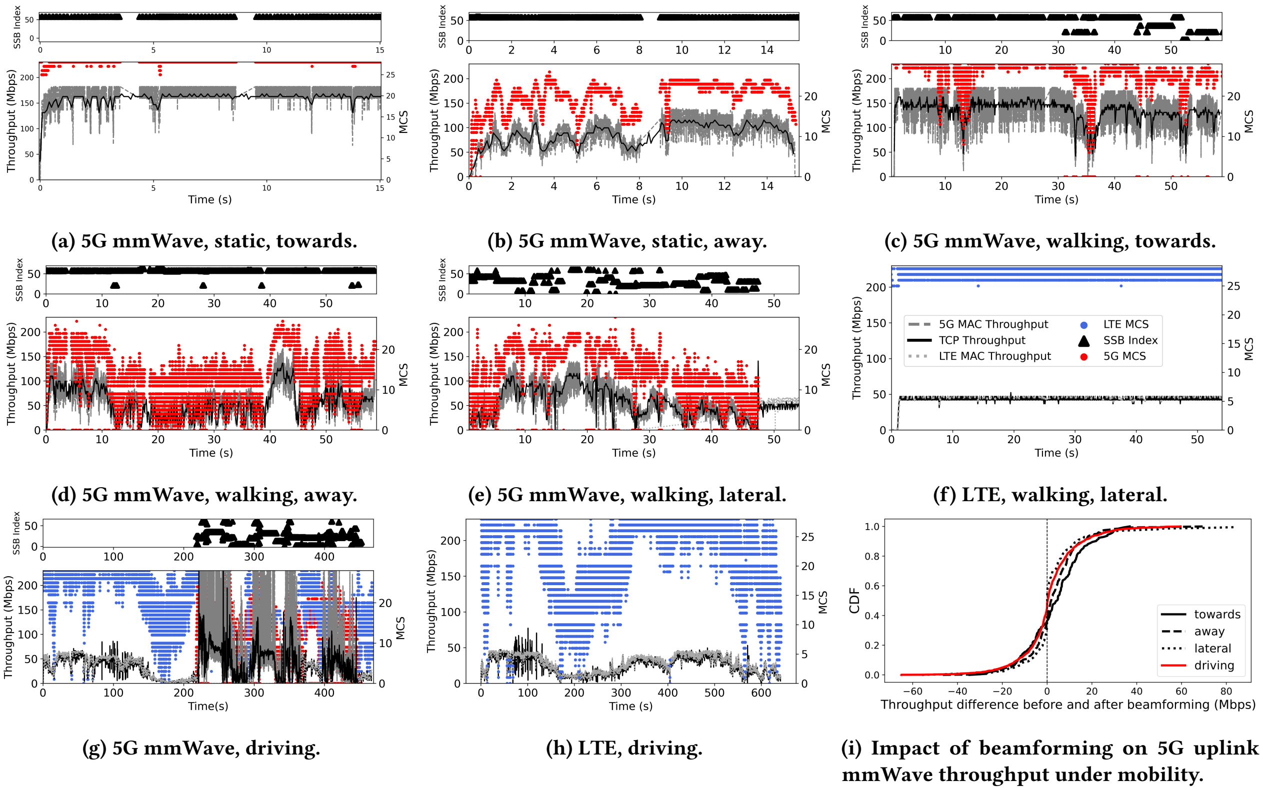
Interestingly, Fig. 7i shows that the beamforming performance during driving is similar to that of walking. This suggests that the beamforming implementation is highly optimized in terms of speed. However, optimizing the beam selection algorithms remains an open challenge.

Dual connectivity. The UE may fall back from 5G mmWave to LTE when the 5G mmWave signal is weak, at the discretion of the BS. In our experiments, this happens occasionally in the laterally walking scenario (Fig. 7e, *e.g.,* from 47 s to 53 s) and very frequently in the driving scenario (Fig. 7g, *e.g.,* from 368 s to 396 s and from 448 s to 468 s). These frequent handovers between LTE and 5G mmWave explain the large latency fluctuations for 5G mmWave during driving in Table 1.

4 ADDITIONAL MEASUREMENTS

Methodology. We conducted additional 5G mmWave measurements in Chicago and Indianapolis. Unlike in Boston, where Verizon is the only operational 5G mmWave carrier, in Chicago and Indianapolis, both Verizon and AT&T (operating only in 39 GHz – n260 band) offer 5G mmWave coverage. However, we found AT&T in Chicago employs a rate-limiting policy after a session of backlogged traffic for the next 15-20 minutes, rendering extensive measurements impossible. Hence, we conducted measurements only with Verizon in Chicago. We also observed that, in contrast to Boston, the mmWave antennas in Chicago and Indianapolis are mounted on the top of traffic lights (Figs. 1b, 1c), about 10-20 ft above the ground in a dense manner, mostly a block away from each other. Our measurement used an AWS Wavelength with Verizon in Chicago and an AWS cloud server with both carriers in Indianapolis, which has no AWS Wavelength deployment.

Results. To understand the baseline 5G mmWave uplink performance, we conducted throughput measurements 50 ft away from the 5G mmWave BS with the UE facing it. The average throughput was 47.21 ± 14.57 Mbps and 43.98 ± 4.83 Mbps in Chicago and Indianapolis, respectively. Compared to Boston (150 Mbps average), the 5G mmWave performs suboptimally in Chicago and Indianapolis (see Fig. 8a for a representative run in each of the three cities). On the other hand, AT&T in Indianapolis yields good average uplink performance (150 ± 50 Mbps) similar to that of Verizon in Boston. However, the AT&T uplink throughput in Indianapolis varies significantly across different runs and sometimes even for the same run, as shown in Runs 1, 2, 3 in Fig. 8b. In Run 1, the average throughput is around 230 Mbps, whereas in Run 2, the throughput is sub-optimal, very close to what we observed for Verizon in Chicago and Indianapolis. Finally, in Run 3, the AT&T throughput oscillates from a high value of 250 Mbps to as low as 0 Mbps within a few seconds. In contrast, the Verizon throughput in Boston is much more stable over time with only minor fluctuations. Overall, we observe that the 5G mmWave performance varies substantially

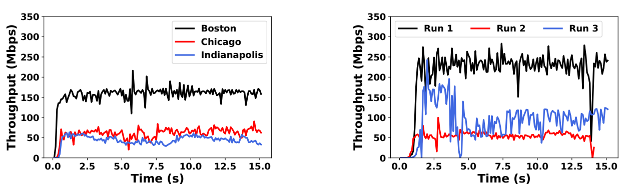


(a) 5G mmWave, static, towards. **(b) 5G mmWave, static, away.** **(c) 5G mmWave, walking, towards.**

(d) 5G mmWave, walking, away. **(e) 5G mmWave, walking, lateral.** **(f) LTE, walking, lateral.**

(g) 5G mmWave, driving. **(h) LTE, driving.** **(i) Impact of beamforming on 5G uplink mmWave throughput under mobility.**

Figure 7: PHY-layer MCS, beam SSB index, MAC-layer throughput and TCP throughput in various scenarios.



(a) Verizon performance in different cities. **(b) AT&T performance in Indianapolis.**

Figure 8: 5G mmWave uplink performance in different cities with different operators.

across operators and across cities for the same operator, as operators employ different resource allocation policies [15].

5 IMPACT ON LCAs

In this section, we make back-of-the-envelope calculations of the expected performance of two classes of edge-assisted LCAs – mobile AR and CAV apps – over today's 5G mmWave deployments, based on the results of our measurement study. For high quality AR that offloads camera frames to an edge server for object detection at 30 or 60 FPS, the end-to-end (E2E) latency (from the moment a frame is captured by the phone's camera to the moment the inference result comes back from the server) should be kept under 33.3 ms or 16.7 ms, respectively. Assuming a standard frame size of 450 KB [25], it would take approximately 37 ms to transfer a frame to an edge server and get back the result in the best case scenario (static, facing the BS) based on the results in Table 1 (158 Mbps bandwidth, 14.5 ms RTT). Similarly, in CAV apps, the end-to-end latency should be kept under 100 ms [23]. Assuming a LiDAR frame size of 2 MB [34], it would take approximately 112 ms to transfer a frame to an edge server and get back the result in the same best case scenario. Additionally, the object detection on the edge server takes time in the order of tens of ms [24, 34], further increasing the E2E latency. These numbers suggest that *today's 5G mmWave with their asymmetric, downlink-centric design and highly fluctuating performance, is far from enabling such LCAs.*

6 NETWORK OPERATOR OPTIMIZATIONS

In this section, we sketch how LCA-awareness could innovate on 5G mmWave designs (which call for architecture/protocol changes) and operations (which can be readily implemented in 5G today) on the UE and infrastructure to overcome the above limitations.

6.1 Elevating Uplink mmWave

Emerging LCAs like edge-assisted AR or CAV apps distinguish themselves with heavy *uplink* data transfers from the UE or cars to infrastructure. This new traffic pattern challenges the traditional wisdom in 5G mmWave, which adopts an *asymmetric* design between uplink and downlink that prioritizes downlink traffic delivery. Hence, an LCA-aware 5G mmWave should explore a paradigm shift from downlink-centric to uplink-centric designs and operations.

Design opportunity: Device-centric uplink mmWave. 5G is infrastructure-centric by design; most mmWave functions are placed at the centralized BSes, leaving limited control to UEs. For uplink-centric apps, this design means that the BS lacks runtime knowledge of uplink traffic and user mobility. Moreover, infrastructure-centric uplink mmWave scheduling incurs heavy signaling overhead, since the BS heavily relies on the UE-side scheduling requests and buffer status reports [7] to initiate uplink scheduling. We believe it is worth offloading more uplink functions from infrastructure to the UE to take advantage of device-side traffic and mobility information.

Operation opportunity: Uplink carrier-aggregation. In the near term, the challenge of accommodating uplink-centric LCAs in the downlink-centric 5G can be mitigated by accelerating the 5G carrier aggregation (CA) deployment at uplink. 3GPP standards [33], 5G operators, and phone vendors currently prioritize downlink CA. For example, the Google Pixel 5 supports 4-CC at downlink, but only 1-CC at uplink and newer phones, such as the Samsung Galaxy S21 Ultra 5G phone, support 8-CC at downlink, but only 2-CC at uplink. We urge 5G infrastructure and UEs to accelerate their support for higher CA at uplink.

6.2 Combating Channel Fluctuations

To stabilize LCA-perceived bandwidth, 5G mmWave should either *reduce* its channel fluctuations, or *mask* fluctuations from upper-layer applications. Both call for extensive efforts in mmWave physical-layer designs and operations.

Design opportunity: QoE-driven mmWave adaptation. Today, most mmWave functions (*e.g.,* rate adaption [4, 5] and scheduling [6])) are agnostic to upper-layer QoE demands and primarily driven by the physical-layer channel feedback, which is inherently fast-varying due to mmWave's high-frequency nature. Such LCA-agnostic physical-layer adaptation does not necessarily result in high QoE in LCAs. To this end, we envision an LCA QoE-driven adaptation may be more effective. This can be made possible by device-centric uplink mmWave in §6.1, the maturity of full-stack reinforcement learning [18], and cross-layer interfaces between applications and cellular stack (*e.g.,* O-RAN [1]).

Operation opportunity: LCA-aware 5G dual connectivity. 5G operators can complement fast-varying mmWave with more stable sub-6 GHz and 4G LTE bands via dual connectivity (DC). With DC, the device can offload the important frames to 4G LTE or sub-6 GHz bands, and leave other frames to mmWave. In practice, this approach can be readily deployed cooperatively by (1) activating DC (by operators); (2) mapping each band to a virtual network interface in the device OS (by phone vendors, similar to having separate voice/data interfaces in VoLTE phones [19]); and (3) splitting frames between different interfaces (by LCAs with domain knowledge). Moreover, it would be also interesting to explore multi-path TCP on 5G DC to refine LCA experience [13].

6.3 Combating Blockage and Mobility

To cope with blockage and mobility, it is worth exploring the following opportunities in designs and operations: **Design opportunity: Blockage and mobility mitigation by prediction.** The UE can actively sense or predict blockage and changes in its moving direction using the device's built-in sensors [16] and runtime signaling messages [20] or by leveraging out-of-band information[31], ML [9], or camera vision[10]. With this knowledge, the UE can ask the BS to update beams or schedule more radio resources before blockage, avoid unnecessary beam changes, and notify LCAs to invoke early adaptation.

Operation opportunity: Dense cell deployment. To mitigate blockage- and mobility-incurred throughput fluctuation, 5G operators can deploy denser mmWave cells, so that a user can always be served by at least one mmWave cell. Since such deployments can be costly, it is interesting to study cell deployment strategies to balance mmWave performance and costs. An alternative, potentially more affordable approach, is to deploy reconfigurable intelligent surfaces (RISs), *e.g.,* [8, 11, 12, 22, 32], which can dynamically control the direction of the incident signal and are considered as one of the key enablers of the upcoming 6G technology.

6.4 Shortening mmWave Latency

Beyond throughput, latency is also vital for LCA QoE. Although the 5G mmWave latency is already significantly lower than the LTE latency, there are opportunities to further reduce LCA-perceived latency.

Design opportunity: QoE-aware URLLC. 5G promises ultra reliable low latency communications (URLLC) as one of its "killer" usage scenarios. Current 5G NR standards [**?**] prioritize *small* packet transmissions as their first rollout for URLLC. Some optimizations in this category, such as piggybacking small data transmission over signaling channels, cannot ensure low latency for high-volume LCAs. Generic application-agnostic optimizations in 5G (*e.g.,* traffic scheduling by priority) are reaching their limit for further latency reduction. Hence, it is interesting to explore if QoE-aware 5G URLLC can enable low-latency mmWave for LCAs.

Operation opportunity: Proactive scheduling. In 5G, the BS depends on the client-side scheduling request and buffer status feedback to schedule the uplink radio grants in the next round, thus incurring additional round trip delay. Instead, the BS can predict the LCA traffic, and proactively allocate grants *before* traffic arrives at the UE-side buffer, thus accelerating the uplink transfer since the LCA traffic would no longer have to wait for grants in scheduling.

7 CONCLUSION

We conducted an in-depth measurement study of 5G mmWave uplink performance in 3 major US cities and across 2 mobile operators. Our findings revealed that the uplink 5G mmWave performance exhibits significant diversity across operators and across cities for the same operator. While 5G mmWave provides a 3x higher throughput and 3x lower latency compared to 4G LTE when the UE faces the BS, these gains are not always enough to support emerging LCAs, such as high-quality AR or CAV apps. More importantly, its performance is often suboptimal with a high level of fluctuations under challenging scenarios, such as self-blockage or mobility, which can further degrade the performance of LCAs. Our detailed analysis of 5G specific signalling messages and PHY-level KPIs showed that that the suboptimal performance cannot be easily fixed via simple tuning of network configurations. Consequently, we identified a comprehensive list of design and deployment optimizations that 5G operators can explore to bring 5G mmWave performance to the level needed to ultimately support the LCAs.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their helpful comments. This work was supported in part by NSF grant 2112778-CNS. Yuanjie Li acknowledges support from the National Key Research and Development Plan of China (2018YFB1800301) and the National Natural Science Foundation of China (61832013).

REFERENCES

[1] 2021. O-RAN Alliance: Operator Defined Open and Intelligent Radio Access Networks. https://https://www.o-ran.org/.

[2] Online. *AWS Wavelength - Deliver ultra-low latency applications for 5G devices.* https://aws.amazon.com/wavelength/

[3] Online. *nuttcp - Network Performance Measurement Tool.* https://www.nuttcp.net

[4] 3GPP. 2019. TS38.211: 5G NR; Physical channels and modulation.

[5] 3GPP. 2019. TS38.213: 5G NR; Physical layer procedures for control.

[6] 3GPP. 2019. TS38.321: 5G NR; Medium Access Control (MAC) protocol specification.

[7] 3GPP. 2021. TS38.331: 5G NR: Radio Resource Control (RRC).

[8] Omid Abari, Dinesh Bharadia, Austin Duffield, and Dina Katabi. 2017. Enabling High-Quality Untethered Virtual Reality. In *Proc. of USENIX NSDI.*

[9] Ahmed Alkhateeb and Iz Beltagy. 2018. Machine Learning for Reliable mmWave Systems: Blockage Prediction and Proactive Handoff. In *Proc. of IEEE GlobalSIP.*

[10] Muhammad Alrabeiah, Andrew Hredzak, and Ahmed Alkhateeb. 2020. Millimeter Wave Base Stations with Cameras: Vision-Aided Beam and Blockage Prediction. In *Proc. of IEEE VTC-Spring.*

[11] Ertugrul Basar. 2020. Reconfigurable Intelligent Surface-Based Index Modulation: A New Beyond MIMO Paradigm for 6G. arXiv:1904.06704v3

[12] Ertugrul Basar, Marco Di Renzo, Julien De Rosny, Merouane Debbah, Mohamed-Slim Alouini, and Rui Zhang. 2019. Wireless Communications Through Reconfigurable Intelligent Surfaces. *IEEE Access* 7 (2019), 116753–116773.

[13] CableLabs. 2019. 5G Link Aggregation with Multipath TCP (MPTCP) . https://www.cablelabs.com/5g-link-aggregation-mptcp.

[14] Paolo Collela. 2017. *5G and IoT: Ushering in a new era.* https://www.livemint.com/Opinion/SktcUSRU6iMQ7BNUknwbFK/5G-and-IoT-Ushering-in-a-new-era.html

[15] Phuc Dinh*, Moinak Ghoshal*, Dimitrios Koutsonikolas, and Joerg Widmer. 2022. Demystifying Resource Allocation Policies inOperational 5G mmWave Networks. In *2022 IEEE 22nd International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM).*

[16] Zhihan Fang, Guang Wang, Xiaoyang Xie, Fan Zhang, and Desheng Zhang. 2021. Urban Map Inference by Pervasive Vehicular Sensing Systems with Complementary Mobility. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (2021), 1–24.

[17] IEEE 802.11 Working Group. 2012. IEEE 802.11ad, Amendment 3: Enhancements for Very High Throughput in the 60 GHz Band. (2012).

[18] Samuel Joseph, Rakesh Misra, and Sachin Katti. 2019. Towards self-driving radios: Physical-layer control using deep reinforcement learning. In *Proceedings of the 20th International Workshop on Mobile Computing Systems and Applications.* 69–74.

[19] Chi-Yu Li, Guan-Hua Tu, Chunyi Peng, Zengwen Yuan, Yuanjie Li, Songwu Lu, and Xinbing Wang. 2015. Insecurity of voice solution volte in lte mobile networks. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security.* 316–327.

[20] Yuanjie Li, Chunyi Peng, Zengwen Yuan, Jiayao Li, Haotian Deng, and Tao Wang. 2016. Mobileinsight: Extracting and Analyzing Cellular Network Information on Smartphones. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking* (New York City, New York) *(MobiCom '16).* Association for Computing Machinery, New York, NY, USA, 202–215. https://doi.org/10.1145/2973750.2973751

[21] Yuanjie Li, Chunyi Peng, Zhehui Zhang, Zhaowei Tan, Haotian Deng, Jinghao Zhao, Qianru Li, Yunqi Guo, Kai Ling, Boyan Ding, et al. 2021. Experience: a five-year retrospective of MobileInsight. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking.* 28–41.

[22] Christos Liaskos, Shuai Nie, Ageliki Tsioliaridou, Andreas Pitsillides, Sotiris Ioannidis, and Ian Akyildiz. 2020. End-to-End Wireless Path Deployment With Intelligent Surfaces Using Interpretable Neural Networks. *IEEE Transactions on Communications* 68 (2020), 6792–6806. Issue 11.

[23] Shih-Chieh Lin, Yunqi Zhang, Chang-Hong Hsu, Matt Skach, Md E Haque, Lingjia Tang, and Jason Mars. 2018. The architectural implications of autonomous driving: Constraints and acceleration. In *Proc. of ACM ASPLOS.*

[24] Luyang Liu, Hongyu Li, and Marco Gruteser. 2019. Edge assisted real-time object detection for mobile augmented reality. In *The 25th Annual International Conference on Mobile Computing and Networking.* 1–16.

[25] Jiayi Meng, Z. Jonny Kong, Y. Charlie Hu, Mun Gi Choi, and Dhananjay Lal. 2022. Do We Need Sophisticated System Design for Edge-assisted Augmented Reality?. In *Proc. of ACM EdgeSys.*

[26] Arvind Narayanan, Eman Ramadan, Jason Carpenter, Qingxu Liu, Yu Liu, Feng Qian, and Zhi-Li Zhang. 2020. A First Look at Commercial 5G Performance on Smartphones. In *Proc. of ACM WWW.*

[27] Arvind Narayanan, Muhammad Iqbal Rochman, Ahmad Hassan, Bariq S. Firmansyah, Vanlin Sathya, Monisha Ghosh, Feng Qian, and Zhi-Li Zhang. 2022. A Comparative Measurement Study of Commercial 5G mmWave Deployments. In *Proc. of IEEE INFOCOM.*

[28] Arvind Narayanan*, Xumiao Zhang*, Ruiyang Zhu, Ahmad Hassan, Shuowei Jin, Xiao Zhu, Denis Rybkin, Dustin Zhang, Michael Yang, Z. Morley Mao, Feng Qian, and Zhi-Li Zhang. 2021. A Variegated Look at 5G in the Wild: Performance, Power, and QoE Implications. In *Proc. of the ACM Sigcomm.*

[29] Muhammad Iqbal Rochman, Vanlin Sathya, Norlen Nunez, Damian Fernandez, Monisha Ghosh, Ahmed S. Ibrahim, and William Payne. 2022. A Comparison Study of Cellular Deployments in Chicago and Miami Using Apps on Smartphones. In *Proceedings of the 15th ACM Workshop on Wireless Network Testbeds, Experimental Evaluation, CHaracterization.* Association for Computing Machinery, New York, NY, USA.

[30] Stefano Sorrentino. 2021. *Powering connected cars with 5G.* https://www.ericsson.com/en/blog/2021/10/powering-connected-cars-with-5g

[31] Sanjib Sur, Ioannis Pefkianakis, Xinyu Zhang, and Kyu-Han Kim. 2017. WiFi-Assisted 60 GHz Wireless Networks. In *Proc. of ACM MobiCom.*

[32] Xin Tan, Zhi Sun, Dimitrios Koutsonikolas, and Josep Miquel Jornet. 2018. Enabling Indoor Mobile Millimeter-wave Networks Based on Smart Reflect-arrays. In *Proc. of IEEE INFOCOM.*

[33] Techplayon. 2018. 3GPP Release-15 UE Cat Types, Throughput, Modulation and DL / UL Combinations. https://www.techplayon.com/3gpp-release-15-ue-category-types-throughput-performance-modulation-and-supported-ue-dl-ul-combinations/.

[34] Xumiao Zhang, Anlan Zhang, Jiachen Sun, Xiao Zhu, Y. Ethan Guo, Feng Qian, and Z. Morley Mao. 2021. EMP: Edge-Assisted Multi-Vehicle Perception. In *Proc. of ACM Mobicom.* Association for Computing Machinery.