A Variegated Look at 5G in the Wild: Performance, Power, and QoE Implications

Arvind Narayanan[†*], Xumiao Zhang[‡*], Ruiyang Zhu[‡], Ahmad Hassan[†], Shuowei Jin[‡], Xiao Zhu[‡],
Xiaoxuan Zhang[†], Denis Rybkin[†], Zhengxuan Yang[†], Z. Morley Mao[‡], Feng Qian[†], Zhi-Li Zhang[†]

[†]University of Minnesota – Twin Cities   [‡]University of Michigan – Ann Arbor

ABSTRACT

Motivated by the rapid deployment of 5G, we carry out an in-depth measurement study of the performance, power consumption, and application quality-of-experience (QoE) of commercial 5G networks in the wild. We examine different 5G carriers, deployment schemes (Non-Standalone, NSA vs. Standalone, SA), radio bands (mmWave and sub 6-GHz), protocol configurations (*e.g.,* Radio Resource Control state transitions), mobility patterns (stationary, walking, driving), client devices (*i.e.,* User Equipment), and upper-layer applications (file download, video streaming, and web browsing). Our findings reveal key characteristics of commercial 5G in terms of throughput, latency, handover behaviors, radio state transitions, and radio power consumption under the above diverse scenarios, with detailed comparisons to 4G/LTE networks. Furthermore, our study provides key insights into how upper-layer applications should best utilize 5G by balancing the critical tradeoff between performance and energy consumption, as well as by taking into account the availability of both network and computation resources. We have released the datasets and tools of our study at `https://github.com/SIGCOMM21-5G/artifact`.

CCS CONCEPTS

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

KEYWORDS

5G, mmWave, Network Measurement, Power Model, Power Characteristics, Energy Efficiency, Latency, Video Streaming, Dataset

---

* These authors contributed equally to this paper.
Corresponding authors: arvind@cs.umn.edu, xumiao@umich.edu.

---

1 INTRODUCTION

5G New Radio (NR) specifications [20] open a wide spectrum of frequencies. High-band millimeter wave (mmWave) 5G, along with its mid-/low-band sub-6 GHz counterpart, make up the current 5G market. We pay close attention to mmWave 5G due to its ultra-high bandwidth which attracts emerging bandwidth-hungry applications. On the other hand, mmWave is very sensitive to factors such as mobility and blockage due to its much shorter wavelength, making the upper-layer network management (*e.g.,* bitrate adaptation of video streaming) more challenging. Despite numerous studies on modeling and simulation of mmWave links [27, 29, 34, 35, 50, 67, 68, 70], the impact of mmWave on commercial 5G performance, power consumption, as well as mobile application Quality-of-Experience (QoE) is largely under-explored.

In addition to its high bandwidth and low latency enabled by physical-layer innovations (*e.g.,* massive MIMO, advanced channel coding, *etc.*), power saving is a top concern to mobile users of 5G. In cellular networks, this is usually achieved by different Radio Resource Control (RRC) states. 5G makes no exception. It is thus important to understand the RRC state machine of commercial 5G networks and its implications. To reduce time to market, most carriers employ the Non-Standalone (NSA) mode for their initial deployment. NSA leverages 5G for data plane operations while reusing the existing 4G infrastructure for control plane operations, making the RRC state machine 4G-like. Very recently, Standalone (SA) 5G deployment has hit the commercial landscape. SA is completely independent of the legacy 4G cellular infrastructure, fully unleashing the potential of 5G. The configurations of key parameters in the state machine lead to important performance and energy trade-offs. They are usually carrier-specific and can be very different between NSA and SA deployment modes.

In order to understand commercial 5G networks' end-to-end performance and power characteristics, as well as their Quality of Experience (QoE) implications on mobile applications, in this paper, we conduct a comprehensive yet in-depth measurement study of two commercial 5G networks in the US. As 5G technology evolves, its performance is expected to improve over time. We therefore compare our measurement results with earlier studies to get the initial longitudinal insights on 5G's evolution. We also compare our findings on mmWave with its low-band counterpart. Our study faced a number of challenges:

• 5G-NR supports a wide range of frequency spectrum: low-band, mid-band, and mmWave. All these frequency bands have different performance and signal propagation characteristics. Additionally, 5G can be deployed in either SA or NSA mode, which further has implications on performance [30]. Conducting a measurement study on such a heterogeneous ecosystem is challenging.

Arvind Narayanan*, Xumiao Zhang*, Ruiyang Zhu, Ahmad Hassan, Shuowei Jin, *et al.*

- The coverage of different bands and deployment modes is often sporadic. For instance, in the case of mmWave with poor signal propagation characteristics, most of its deployment is outdoor. Surveying the availability of band-specific 5G service requires extensive field experiments.
- Evaluating mobile carriers' end-to-end network performance in the wild is known to be difficult. Many entities can become the performance bottleneck including the Internet, mobile carrier's infrastructure, as well as end devices themselves. Identifying the bottleneck in mmWave 5G is particularly challenging due to its ultra-high bandwidth.
- 5G power measurement is not trivial. The state-of-the-art hardware power monitors often require a stable external power supply, making mobility experiments difficult to perform in the wild. In addition, vendors have been making smartphones more "closed" by integrating the battery and back cover with the main body. Skilled engineering efforts are required to connect off-the-shelf 5G smartphones to a power monitor. It's not easy to connect any commodity off-the-shelf (COTS) 5G capable smartphones to a power monitor and conduct outdoor experiments.
- To understand the benefits that 5G brings to mobile applications and to identify the new challenges in 5G, we need fair comparisons with the 4G baseline. However, 4G and 5G have very different characteristics, making it difficult to experimentally compare them in a fair, efficient, and representative way.

To address these challenges, we first build a holistic testbed consisting of commercial 5G smartphones, external power monitors, and cloud servers. We further develop a set of software and hardware tools to control the workloads and physical environments, as well as to log important information at different layers in a fine-grained manner. Through carefully designed experiments, we demystify the current 5G performance, power, and QoE implications with special emphasis on mmWave. Our experiments over a 4-month period consumed more than 15 TB of cellular data. The key contributions of our study are summarized as follows.

- We perform a detailed performance examination of 5G over multiple frequency bands including sub-6 GHz and mmWave. We find that both their throughput and latency have experienced noticeable improvements compared to its initial deployment. The end-to-end performance is highly correlated with geographical properties. We quantify such properties and their vastly different impacts on NSA and SA 5G. In particular, we perform experiments over T-Mobile's SA 5G deployed for their low-band network. This is to our knowledge the first examination of commercial SA 5G performance.
- Through principled probing algorithms, we infer the RRC states and configuration parameters for SA 5G (T-Mobile) and NSA 5G (Verizon and T-Mobile). For NSA 5G that relies on 4G as an anchor, we find that the NR_RRC_CONNECTED to LTE_RRC_IDLE state transition (due to data inactivity on UE) for the carriers considered in our study is 2× more energy efficient than those studied in a previous NSA 5G measurement study [59].
- We take a closer look at the power characteristics of 5G and 4G/LTE. Over downlink (uplink), 5G can be 79% (74%) less energy-efficient than 4G at low throughput but up to 5× (2×) more energy-efficient when the throughput is high. Using a data-driven approach, we build a *first* throughput and signal strength-aware radio power model for different frequency bands of 5G.

Table 1: Statistics of the data collected using two commercial 5G carriers: Verizon and T-Mobile.

Dataset Statistics	
5G Network Performance Tests	12,500+
Unique servers tested with	157+
Cumulative time of measurement traces	2,666 minutes+
Power Measurements @ 5000 Hz	2,336 minutes+
Total kilometers walked	148.5 km+
# of real Web Page Load Tests	30,000+
# of 5G smartphones (and models)	7 (3)

- We conduct a *first* evaluation of state-of-the-art adaptive video bitrate adaptation (ABR) algorithms over mmWave 5G, which is the key radio technology for supporting ultra-high definition (UHD) videos and beyond. We find that due to the poor signal propagation characteristics of mmWave 5G, existing ABR mechanisms over mmWave 5G can incur ∼3.7% to 259.5% higher stall time than 4G/LTE. We propose simple yet effective interface selection mechanisms for 5G video streaming. It can yield a 26.9% video stall reduction and a 4.2% improvement in energy efficiency without compromising user-perceived video quality, compared to unmodified streaming algorithms.
- We collect a large dataset consisting of more than 30,000 web page loadings of diverse websites, and use it to compare mmWave 5G vs. 4G page load time and energy consumption. We find that overall 5G improves the page load time at the cost of higher energy consumption compared to 4G. Moreover, this impact is highly web-page-dependent. We build decision tree models that can intelligently select the appropriate network (5G or 4G) for web browsing.
- We have released the functional artifacts (both datasets and tools) of our study: https://github.com/SIGCOMM21-5G/artifact.

Ethical Considerations. This study was carried out by paid and volunteer students. No personally identifiable information (PII) was collected or used, nor were any human subjects involved. We purchased multiple unlimited cellular data plans from Verizon and T-Mobile. Our study complies with the wireless carriers' customer agreements. This work does not raise any ethical issues.

2 MEASUREMENT SETTINGS & TOOLS

5G Carriers, 5G Bands and Locations. Since its commercial launch, the 5G ecosystem – which includes service deployments, coverage, 5G-capable devices – is rapidly expanding and evolving. In our measurement study, we select two commercial carriers in the US for our experiments – Verizon and T-Mobile. While both these carriers have deployed 5G services on several bands, in our dataset, we find that Verizon has deployed NSA-based 5G service that provides both mmWave 5G over 28/39 GHz frequency bands (n261/n260) and low-band 5G (n5) w/ 4G bands by leveraging dynamic spectrum sharing (DSS) technology. In contrast, T-Mobile provides low-band (@ 600MHz or n71)[1] 5G service using *both NSA and SA* modes. The measurement study is conducted in two US cities where both carriers have deployed 5G services. Key statistics of the datasets collected are summarized in Table 1.

5G UE and Android Measurement Tool. We use multiple smartphone models of user equipment (UE) with 5G support: Google

---

[1]T-Mobile also provides NSA-based mid-band (n41) and mmWave 5G (n261/n260) in select areas, however these services were not considered in this study.

Pixel 5 (PX5), Samsung Galaxy S20 Ultra 5G (S20U) and Samsung Galaxy S10 5G (S10). These phones have diverse specifications. For instance, compared to PX5, S20U has a superior chipset, 5G modem, increased RAM and CPU frequencies. We make considerable additions to 5G Tracker [41] and build a comprehensive monitoring toolkit with various functions to monitor network traffic, battery status (current and voltage), signal strength *etc.*. Some of these functions require rooting the phones. We use both rooted and non-rooted phones (based on needs) to measure various aspects of 5G performance and power usage under different settings.

Power Monitoring Tool. We use Monsoon Power Monitor [17] to power smartphones and measure the power consumption. For outdoor walking experiments, we use a portable external power source to supply power to the monitor.

3 IMPROVEMENTS AND NEW FINDINGS IN 5G NETWORK PERFORMANCE

In this section, we closely examine the end-to-end network performance of commercial 5G networks by conducting several carefully designed experiments in the wild.

3.1 Measurement Methodology

Challenges. There are several known challenges while evaluating end-to-end network performance of mobile carriers in the wild. **[C1]** First, Internet-side congestion can adversely affect network performance. **[C2]** Secondly, we also have no clear visibility into the carrier's network/transport infrastructure and policies enforced by them. **[C3]** Finally, there is significant diversity in end-device (*e.g.,* server or smartphone) specifications and capabilities which can affect network performance.

Methodology. We now describe our carefully designed methodology for evaluating 5G network performance. Ookla's *Speedtest* service [43] is a widely used and state-of-the-art tool for testing Internet connection bandwidth and latency. By default, Speedtest chooses a geographically nearby server with the least round-trip latency to measure downlink/uplink throughput. They also allow users to choose a server from a pool of geographically distributed servers. More importantly, both the 5G carriers studied host servers on Ookla. For instance, Verizon hosts 48 servers while T-Mobile hosts 47 servers. These are mainly located in major metropolitan U.S. cities. We leverage the flexibility of server selection as well as the carrier's presence in Speedtest's pool of server network to evaluate a carrier's network performance by conducting several tests on carrier-hosted servers. Particularly, this strategy helps us reduce the impact of **[C1]** and **[C2]** on our measurement tests.

The default policy of server selection from Speedtest is to choose a server located in the same city as the UE. Our results (Fig. 24 in Appendix A.2) also further confirm that using carrier-hosted Speedtest servers (especially if one is available in the UE's city) usually provides best performance over non-carrier based servers. Even when testing using carrier-hosted servers in other states and cities, we believe this strategy helps eliminate most of the Internet side bottleneck as the carrier would usually place Speedtest servers at the edge of the carrier's city-level ingress points. Speedtest service uses TCP protocol for all its tests. Speedtest additionally also allows us to conduct a test in one of the two connection modes: **(i)** using a single connection and **(ii)** using multiple connections that
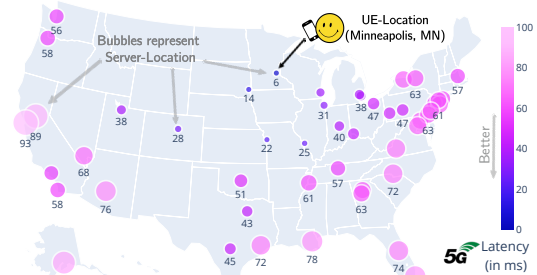

Figure 1: Impact of UE-Server distance on RTT.

is non-configurable. The number of multiple connections varies from one test to another, and the algorithm is not disclosed on how Speedtest decides the number of connections to establish for a test. To account for this limitation, we also provision VMs with high network-throughput (in different U.S. locations) provided by Microsoft Azure's public cloud service. This allowed us to evaluate the impact of different transport layer protocols and parameters.

Lastly, we take two steps to address **[C3]**. First, to account for UE diversity, we use two 5G smartphones: PX5 and a more powerful S20U (§2). Secondly, in addition to the carrier-hosted Speedtest servers, we also use *all* the Speedtest servers located in the local state of the UE. This allows to reduce the impact of geographic distance on network performance, rather allows us to understand the impact of other *potential* server-side factors over 5G network performance. For each unique `<UE-model,carrier,server>` setting, we repeat the test at least 10 times per connection mode. Our dataset contains over 12,500 Speedtest measurements[2]. We report the 95th percentile performance results of all Speedtest sessions repeatedly conducted for a setting. In other words, our approach measures the peak network performance, and should *not* be confused with the user perceived network quality metrics [5]. Focusing on the peak metrics helps us to further reduce the impact of congestion and other Internet-side factors on our performance measurements, and rather helps us understand the impact of UE-Server distance and radio technology/band over network performance. Having this information is particularly important for application and service providers so that they can better harness 5G. Unless specified otherwise, all mmWave-5G based experiments were conducted outdoors and the UE was held stationary with clear LoS to the 5G tower.

Baseline. To provide the initial longitudinal insights of commercial 5G's network performance in the US, we consider *5Gophers* [39] dataset (reportedly measured in the US as of October 2019) as the baseline for comparing results.

3.2 Impact of UE-Server Distance

Round-Trip Time (RTT). By tapping into the 5G carriers' nationwide network of Speedtest servers, we next quantify the impact of UE-Server distance over RTT. Fig. 1 shows the latency characteristics of Verizon's mmWave 5G service for different server locations on a geographic map. UE's location is fixed as Minneapolis, MN. Clearly, RTT degrades severely as the UE-Server distance increases. The lowest observed RTT is ~6ms when tested with a server located closest (~3 km) to the UE. Compared to latency observed back in 2019 [39] (*i.e.,* during early deployment), this is a ~50% improvement

---

[2]We developed scripts for Android smartphones to completely automate the process of conducting a test using Ookla's Speedtest service (free version).
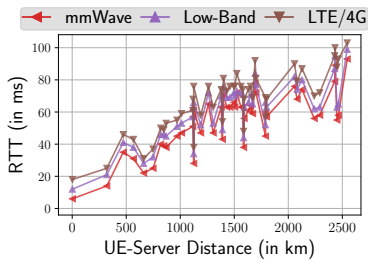
Arvind Narayanan*, Xumiao Zhang*, Ruiyang Zhu, Ahmad Hassan, Shuowei Jin, *et al.*

Figure 2: [Verizon] Latency.

Figure 3: [Verizon] Downlink.

Figure 4: [Verizon] Uplink.

Figure 5: [T-Mobile] Latency.

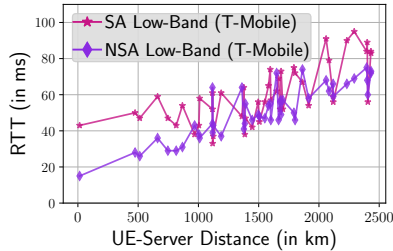**Figure 6: [T-Mobile] Downlink.**

Figure 7: [T-Mobile] Uplink.

over the baseline performance. RTT gets doubled as the UE-Server distance increases to 320 km. This trend is more clearly visible in Fig. 2[3] which further signifies the importance of *edge computing* for latency-sensitive apps and services. Fig. 2 also compares RTT values of mmWave 5G against that of low-band 5G and 4G/LTE. We find that low-band 5G suffers an additional delay of ~6 to 8 ms over mmWave 5G across the entire UE-Server distance range. This is not surprising as mmWave 5G bands (n260/n261) with higher sub-carrier spacing and shorter OFDM symbol duration lead to lower latency when compared to low-band 5G [53, 54]. On the other hand, due to flexible frame structure and fine-grained transmission time interval (TTI) in 5G NR, we find both low-band and mmWave 5G exhibit better RTT (6 to 15ms reduction) than LTE. Similar experiments were also conducted over T-Mobile's network (including SA Low-Band 5G) and results are shown in Fig. 5. While the earlier trend observed in Verizon's network about the impact of UE-Server distance over RTT also holds true for T-Mobile's network, we do not find any significant difference yet in RTT performance between T-Mobile's SA and NSA deployments of low-band 5G.

Throughput. Fig. 3 shows the impact of UE-Server distance on Verizon mmWave 5G downlink throughput performance. With multiple TCP connections, the UE is able to achieve an impressive downlink throughput of over 3 Gbps across all the servers in the US. This is a ~50-60% improvement over the baseline. We attribute this improvement to ramping up of carrier aggregation from 4CC to 8CC which requires improvements in carrier's infrastructure as well as the UE's chipset specifications (see Appendix A.1). As pointed out earlier, Speedtest does not allow us to control the number of TCP connections for a test. Using packet dumps, we found that Speedtest would establish anywhere between 15 to 25 TCP connections for the multiple connection test. The packet loss rate was less than 1%. However, with a single TCP connection, we find that the throughput degrades as the UE-Server distance increases (see Fig. 3). We suspect this degradation is due to the: **(1)** increase in RTT which is known to affect TCP performance, **(2)** packet loss
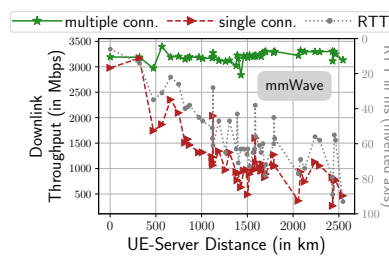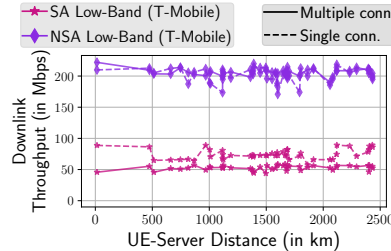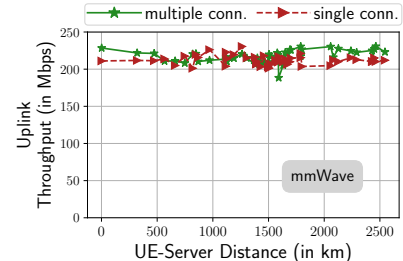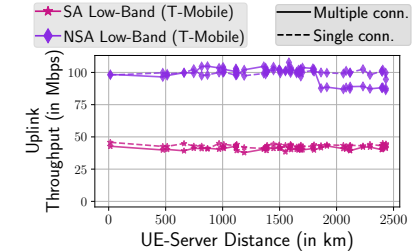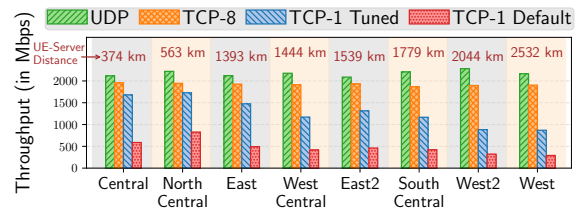

Figure 8: Single conn. downlink throughput across all US-based Azure regions under different transport layer settings.

(even at the slightest rate). The impact of both coupled with existing TCP mechanisms gets amplified at ultra-high bandwidth levels thus degrading TCP performance. Nonetheless, compared to the baseline, we find there is a significant improvement in the single TCP (1-TCP) connection's performance. 1-TCP connection (with less overhead compared to multiple connections) can also achieve close to 3 Gbps throughput provided the server is much closer to the UE. This again signifies the importance of the edge especially for bandwidth-hungry applications. Uplink throughput (see Fig. 4) performance has also improved by a factor of 3× to 4× over the baseline. Both single and multiple connection uplink tests can achieve a throughput of ~220 Mbps. On the other hand, for T-Mobile which also has SA-based deployments for the low-band 5G, we find that both downlink and uplink performance can achieve only half the performance of what their low-band NSA 5G service can provide (see Figs. 6 and 7). We believe this to be due to carrier aggregation not yet supported for SA or that the 5G core is not fully mature to provide the benefits envisioned by SA 5G.

Taking a Closer Look at Single-Conn. Throughput. To get a better understanding of single-connection's performance with mmWave 5G (known to provide ultra-high bandwidth capacity), we perform controlled experiments using Microsoft Azure's public cloud service. We provision a high-network bandwidth capacity VM (Type: DS4_v2) at every region in the US provided by Microsoft Azure. In order to capture packet dumps and have the ability to change kernel parameters, we use *rooted* PX5 to conduct these experiments. Unlike S20U that can achieve a throughput of more than 3 Gbps, PX5 has a maximum observable downlink throughput

---

[3]Figures 2 to 7 shows servers located in the conterminous US region only.
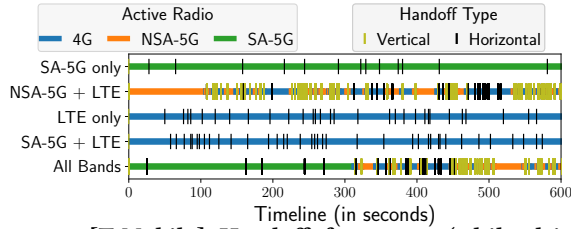
Figure 9: [T-Mobile] Handoff frequency (while driving) across different low-band frequency settings.

of ∼2.2 Gbps (see Appendix A.1 for details). For TCP, we use *CUBIC TCP* [16] as the congestion control algorithm. The experimental setup uses UDP performance as baseline. As shown in the results (see Fig. 8), UDP is able to achieve peak observable throughput across all the server locations. We observe a small yet noticeable gap between UDP and 8-TCP performance most likely due to the protocol overhead of TCP. However, with default Linux kernel (v4.18.0) parameters for TCP, we find 1-TCP connection's throughput is limited to no more than 500 Mbps for all servers. Upon further investigation, increasing the maximum size of TCP write buffer (tcp_wmem) parameter of Linux's TCP kernel significantly improves the UE's downlink throughput using 1-TCP connection by a factor of 2.1× to 3× (denoted as "1-TCP tuned" in Fig. 8). Theoretically, the sender's TCP buffer size (which is a per socket configuration) must at the least be equal to the bandwidth-delay product (BDP) of the high-throughput flow's capacity. In other words, transport-layer kernel parameters should be carefully tuned to meet the desired application QoE requirements. Nonetheless, even the tuned 1-TCP performance falls short by ∼886 Mbps on average when compared to UDP. Similar to the impact of UE-Server distance observed earlier in Fig. 3 for the single-connection performance using Ookla's Speedtest service, we make similar observations in performance under controlled experimental settings using Azure servers. In that, we again find that TCP performance (including that of *1-TCP tuned*) exacerbates as the UE-Server distance increases. These observations highlight the inefficacies that exist in current TCP and congestion control mechanisms over mmWave 5G networks.

3.3 Handoffs in (Low-Band) NSA & SA 5G

Previous studies on handoffs[4] of NSA mmWave 5G [39] have shown that compared to 4G/LTE, there are far more frequent handoffs. This is mainly due to the smaller coverage footprint of mmWave towers as well as that NSA 5G still relies on LTE for control plane signaling. In this preliminary study, we however focus on comparing T-Mobile's SA 5G with NSA 5G that are commercially deployed. T-Mobile is the only carrier that has deployed both NSA and SA-based 5G for their low-band network. To obtain connectivity to SA 5G (over n71 band), it was critical to use T-Mobile's firmware in S20U. We selected a 10 km driving route which traversed via busy downtown regions and freeways with driving speeds ranging from 0 to 100 kph. Using Samsung's service code (*#2263#), we selectively enable a set of radio bands to configure the UE in one of the 5 setting: **(i)** enable SA-n71 band only, **(ii)** enable NSA-n71 and LTE bands only, **(iii)** enable LTE bands only, **(iv)** enable SA-n71 and LTE bands only, and, **(v)** enables all bands (default setting). For each configuration, while the UE was handheld by a passenger, we

---

[4]Handoff here refers to the change in tower or data transmission technology.

drove the route 2× per direction and monitored the handoff activity. Fig. 9 shows a representative set of results. There are five horizontal bars, one for each of the 5 band configuration settings. Within each horizontal bar, there are several colored-segments that denoted the active radio (blue for 4G/LTE, orange for NSA-5G, and green for SA-5G). Ticks on these bars indicate the occurrence of a horizontal handoff (*i.e.,* across towers) or a vertical handoff (*i.e.,* across radio technologies). The most important finding here is that SA 5G has far fewer handoffs (*i.e.,* 13 handoffs) compared to other configurations, NSA-5G + LTE (110), LTE (30), SA+LTE (38) and all bands (64). These will have implications not just on control plane signaling and scheduling overheads, but also over network performance. Due to increased coverage of the low-band RF n71 band, both SA and NSA over n71 band experience very few horizontal handoffs (13 to 20). But, in NSA, we found close to 90 vertical handoffs (*e.g.,* 4G to 5G or vice-versa) highlighting the complexities involved in NSA.

Now that we have seen the network performance characteristics of different 5G technologies, next we investigate how such performance characteristics impact power.

4 POWER CHARACTERISTICS

In this section, we discuss the power characterization of 5G network and compare with the latest 4G results. To better understand the UE's power consumption, we construct power models for different 5G networks with multiple factors including signal strength, throughput, and frequency bands.

4.1 Methodology

RRC state inference. We first derive the built-in radio state machine which was designed for power management of mobile devices, *e.g.,* parameters of RRC states and transitions for 4G [12] and 5G [13]. For the parameter inference, we improve a network-based approach used in [31, 51] to build our own inference tool, RRC-Probe, in which a server sends UDP packets to a client (UE) at different packet intervals and the UE sends an ACK once a packet is received. The length of RTT depends on the UE's instant RRC state when receiving the packet. Therefore, by measuring the RTT for different packet intervals, we can identify different states and calculate the timers for the transitions between states. Note that this approach does not require root access on smartphones.

Power measurement. We use Monsoon power monitor [17] to measure the UE's power consumption for two purposes: First, we aim to understand power consumption during RRC state transitions. To measure this, the UE is left idle without any data activity for sufficient time (20s in our experiments) thus forcing the UE to be in RRC_IDLE state. A server then sends a packet to the UE which subsequently triggers an RRC_IDLE → RRC_CONNECTED transition and switch to 5G. Then the UE starts its inactivity timer and demotes to RRC_IDLE at the end. In this way, the power monitor can capture full tail period[5] for RRC_CONNECTED. Second, to study the throughput-power relationship and its implications on energy efficiency, we control the UE's data transfer throughput while measuring its power. To reduce the impact of power consumption due to display screen and brightness, we set the screen at the maximum

---

[5]The period after Continuous Reception (*i.e.,* when UE finishes its data transfer) and before demoting to RRC_IDLE in which there are discontinuous reception cycles (DRX) and the UE can reduce power consumption.

Arvind Narayanan*, Xumiao Zhang*, Ruiyang Zhu, Ahmad Hassan, Shuowei Jin, *et al.*

brightness level and subtract the screen power (which is obtained separately) from the total when presenting the results. In this study, power (in W) refers to energy consumed per unit time.

Data Collection Methodology. We conduct both controlled and in-the-wild walking experiments to collect network and power traces at two different cities in the US – Minneapolis (MN) and Ann Arbor (MI) – using two commercial 5G carriers (Verizon and T-Mobile). For Verizon, we collect data for their NSA-based mmWave 5G as well as their low-band 5G service. For T-Mobile, we focus on their low-band 5G which is deployed in both SA and NSA modes. For all our experiments, we use two models of 5G smartphones: S10 and S20U. For the walking experiments, we fixed a 20-min loop (~1.6km). While low-band 5G connectivity for both carriers was omnipresent, mmWave was rather limited. The loop contained three mmWave 5G towers each fitted with three directional mmWave transceivers. We collect 10 traces for each unique carrier-mode-band setting (*e.g.,* Verizon-NSA-Low Band). The power monitor collects data at 5000Hz while we set the network logging rate at 10Hz. As the traces are collected separately by 5G Tracker tool [41] and Monsoon power monitor, we synchronize them by starting both loggers at the same time and further verify by correlating measurements activities known to cause significant power jump.

4.2 RRC Parameters and Power

Using RRC-Probe, we infer a list of RRC parameters for 4G and 5G (see Table 7 in Appendix A.3 for details). From the results, we find that the timers of NSA 5G and 4G LTE are very similar. This is because NSA 5G still retains the existing 4G infrastructure for control plane operations while innovating the data plane to enhance the network capacity.

Fig. 10 illustrates the results of inferring the RRC states. For NSA 5G, the RRC states are basically the same as 4G. However, according to the 5G-NR specifications [20], a new RRC state called RRC_INACTIVE is introduced in SA 5G. We believe this new state can be seen in Fig. 10 (see top left part representing T-Mobile SA 5G). We find that the UE remains in this state for about 5s (*i.e.,* 10s to 15s of interval) before transitioning to RRC_IDLE. The main purpose of this state (akin to a low-power state) is to provide an efficient mechanism for the UE's radio to sleep (thus saving power) and at the same time enable a quick and lightweight transition back to the RRC_CONNECTED state (thus improving latency by reducing the radio's wake up time). These benefits are largely achieved by reducing the control plane signaling overhead. Besides, we notice that T-Mobile SA 5G has a tail timer of 10s which is similar to that of T-Mobile NSA 5G and Verizon NSA 5G, indicating UE directly enters RRC_IDLE after leaving RRC_CONNECTED. We also confirm the timers using Monsoon power monitor. This is different from the observations in [59] that found the 5G tail is 20s, *i.e.,* 2× of 4G tail (10s), which indicates the 5G module must go through both 5G and 4G tails before entering RRC_IDLE. Careful attention needs to be given in configuring such timers as they impact energy efficiency.

We next study the impact of 5G on power during RRC state transitions. We calculate the tail power by averaging the power readings during the entire tail period considering both DRX On duration and the rest of the DRX cycle. As shown in Table 2, 5G consumes more energy than 4G during the tail period and for mmWave 5G the tail power is especially higher. This is likely because the UE's radio
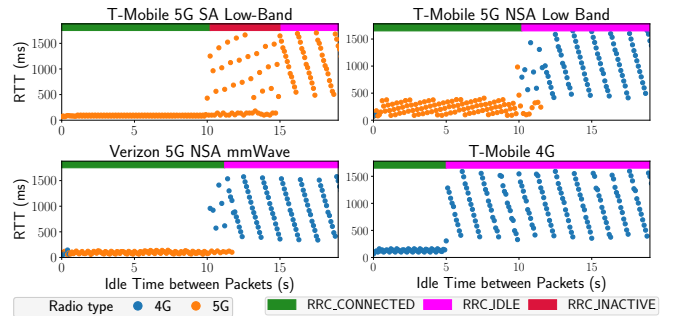

Figure 10: Results of inferring different RRC States using RRC-Probe for SA 5G, NSA 5G and 4G/LTE.

Table 2: Power during RRC state transitions.

Carrier	Network	Power (mW)	
		Tail	*4G→5G switch*
Verizon	4G	178	N/A
T-Mobile	4G	66	N/A
Verizon	NSA 5G (low-band,DSS)	249	799
Verizon	NSA 5G (mmWave)	1092	1494
T-Mobile	NSA 5G (low-band)	260	699
T-Mobile	SA 5G (low-band)	593	245

remains active during the tail period in order to wake up periodically for paging and 5G radio consumes more power than 4G (when the throughput is zero, shown later in §4.3). Further taking into account the 4G → 5G switch power which consumes additional power and is very common (see Fig. 9), 5G is less efficient in terms of state transitions. Therefore, to save power, traffic patterns like periodical data transmission or intermittent waking up should be avoided under 5G. One solution would be forcing the UE to stay in 4G when high throughput is not needed.

4.3 Power for Data Transfer

Previous work on 3G/4G power modeling [31] has constructed power models for data transfer by taking into account the device throughput and concluded that higher throughput leads to higher power consumption. As 5G (*esp.* mmWave) can provide much higher throughput compared to 4G, we study how throughput affects the device power over 5G. With controlled experiments, we measure the device power when transferring data at different download/upload throughput over 4G and 5G. We run UDP data transfer and vary the target throughput using iPerf3. To reduce the impact of poor signal propagation issues of mmWave 5G, we run the experiments by hand-holding the smartphone at a fixed location with Line-of-Sight (LoS) to a 5G panel.

4G vs. 5G. Fig. 11 presents the relationship between throughput and power with a comparison between 4G/LTE and 5G. We also show this relationship across two different bands of 5G: NSA low-band (LB) and NSA mmWave. These experiments were done on S20U[6] over Verizon. We can find that for both 4G and 5G, and for both uplink and downlink directions, the power increases linearly as throughput increases. However, the power for mmWave 5G (uplink and downlink) increases at a slower rate than for the other two radio networks. Although at low throughput levels the power consumption for mmWave 5G is higher, it becomes more efficient when the throughput is high. As seen in Fig. 11, the crossover point

---

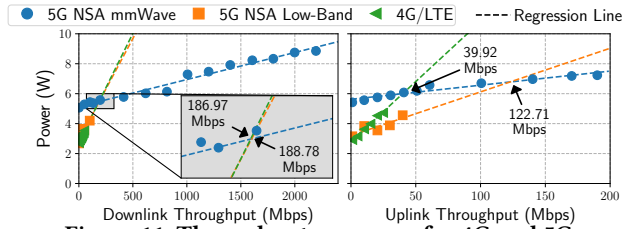[6]Appendix A.4 includes additional results comparing mmWave 5G *vs.* 4G using S10.
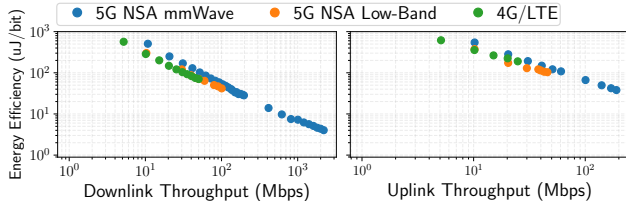
Figure 11: Throughput vs. power for 4G and 5G.

Figure 12: Throughput vs. energy efficiency for 4G and 5G.

Figure 13: Power-RSRP-Throughput relationship.

Figure 14: Energy efficiency-RSRP relationship (mmWave).

at which mmWave 5G becomes more efficient than 4G and low-band 5G is: **(1)** 187 Mbps and 189 Mbps for downlink; **(2)** 40 Mbps and 123 Mbps for uplink. These results clearly reveal the power-performance relationships (and trade-offs) between not just 4G and 5G but also the different bands within 5G. Note that different UE models may have varied levels of power consumption [24]. Similarly, we can observe different crossover points in throughput-power curves between S20U and S10 (see Appendix A.4 for more details). It is also interesting to compare the slopes between low-band 5G and 4G/LTE. In the downlink direction, the slopes of LB-5G ad 4G/LTE are almost identical. In the uplink direction though, LB 5G is much more efficient than 4G/LTE. Next, we calculate the proportion of power consumed by data transfer activity out of total power. On average, data transfer in mmWave 5G consumes 48-76% of the total power consumption for downlink and 46-66% for uplink, while the same for 4G are 21-53% (downlink) and 20-66% (uplink). This is similar to what was also observed earlier by Xu *et al.* [59] (for mid-band 5G). But our results show that the upper bound for 5G downlink is higher by an additional 21% when compared to [59], which is likely due to higher data rates offered by mmWave 5G.

We further calculate the energy efficiency (energy per bit) and plot the results in Fig. 12 with a log scale, where we can also conclude the higher efficiency when transferring at higher speeds under 5G. 5G can be 79% (74%) less efficient than 4G at a low throughput but up to 5× (2×) more when the throughput is high, for downlink (uplink). In fact, this can also be confirmed from mathematical modeling: Assume the device power is P, energy efficiency is E and the throughput is T, we will have $P = c_1 * T + c_2$ and $E = P/T = c_2/T + c_1$. So we can get $\log E \approx c_3 * \log T + c_4$, by taking logarithm on both sides of the equation. Here c_i is constant.

Downlink vs. Uplink. We also compare the downlink transfer with uplink transfer for 4G and 5G (Fig. 11). Based on the carrier configurations, we conclude that the rate of increase in power consumption for uplink is higher by 2.2× to 5.9× than downlink (see Appendix A.4), which is in consensus to prior work on 3G/4G [31].

4.4 Impact of Signal Strength on Power

In addition to throughput, there are other factors affecting the power consumption during data transfer. For example, poor wireless signal strength can negatively affect the device power saving [26, 55]. Moreover, due to poor signal propagation, mmWave's
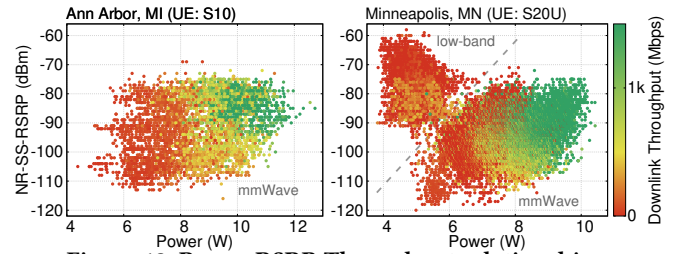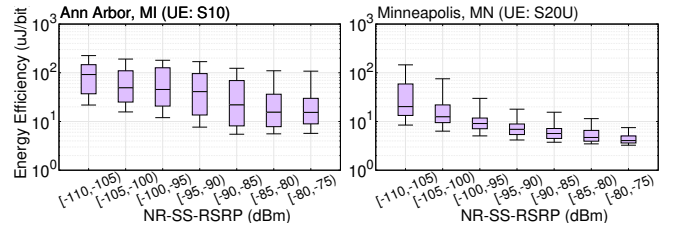
signal strength are known to fluctuate frequently and wildly due to impact of UE-side factors such as mobility or signal reflection characteristics of the surroundings (*e.g.,* open space vs concrete buildings) [40].

We conduct in-the-wild data transfer experiments to collect network throughput and power traces at two locations with Verizon 5G: **(1)** Ann Arbor, MI: mmWave 5G only, **(2)** Minneapolis, MN: both mmWave and low-band 5G. Fig. 13 summarizes how power can be affected by both RSRP and throughput. From the results, we find that **(1)** higher throughput leads to higher power consumption; **(2)** Signal strength also affects the power consumption, which aligns well with earlier findings (§4.3) and previous work [26]. To better isolate the impact of signal strength and understand how it affects power consumption, we show the energy efficiency for different signal strength (RSRP) levels in Fig. 14. As NR-SS-RSRP increases, the energy per bit decreases. This indicates that better signal strength leads to improved energy efficiency. Moreover, for Minneapolis (see right-plot on Fig. 13), we can clearly see there are two clusters of data points. By looking at the network status information, we further confirm that the points in the upper-left cluster represent the data collected when the device is connected to low-band 5G while the other points are for mmWave 5G. In Ann Arbor, we only see mmWave 5G in the logs. Hence, we quantitatively observe that the power consumption varies across different 5G bands that the device is actively using.

4.5 5G Power Model Construction

Previous studies either only consider downlink/uplink throughput [31] or signal strength [24, 42] when modeling the device power for data transfer. However, neither of the assumptions hold given the high variability of 5G throughput in particular for downlink and the vulnerability of 5G signal to the physical environment. Besides, we have seen different bands can have varied power consumption characteristics, hence, it is also important to take into account the band information. To improve model accuracy, we propose to build a network power model for 5G by considering both signal strength and throughput. Based on the observations in §4.3, a linear model can fit well for both uplink and downlink if we solely consider throughput while controlling other factors. However, our
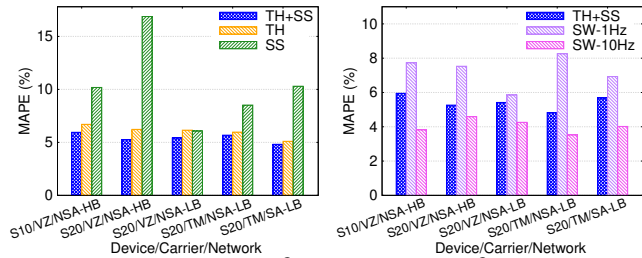
Arvind Narayanan*, Xumiao Zhang*, Ruiyang Zhu, Ahmad Hassan, Shuowei Jin, *et al.*

Figure 15: Comparing performance of different models.

Figure 16: Software monitor calibration.

Table 3: A higher sampling rate incurs more overhead.

Activity	Average Power (mW)
Idle	2014.3
Monitor on (1Hz)	2668.5
Monitor on (10Hz)	3125.7

preliminary experiments show that linearly regressing with multiple factors such as throughput and signal strength together on our walking dataset leads to even higher errors compared to only considering throughput, indicating that the diverse array of multiple impacting factors may not be accurately fit linearly, we instead turn to machine learning-based data-driven approaches to identify the relationships among features for power modeling. Specifically, we apply the Decision Tree Regression (DTR) algorithm.

Model Evaluation. We construct our models and evaluate using a standard metric for regression performance – *Mean Absolute Percentage Error (MAPE)* to reflect the accuracy of our model in terms of relative errors. As observed in §4.4, we construct the power model for different devices (S10, S20U), networks (Verizon, T-Mobile), and radio technologies (NSA/SA, mmWave/low-bands) separately. Note we build models for each setting as opposed to using such information as a feature in the model. We also generate models using previous approaches for comparison. We plot the performance results for all the models in Fig. 15, in which **TH+SS** represents our model which takes into account both throughput and signal strength while **TH** and **SS** represent the models generated only considering throughput or signal strength, respectively. Our models always outperform the models generated from both the previous approaches, which indicates that both features play an important role in affecting the device network power consumption. Without considering the throughput information, the errors of **SS** models are found to be huge compared to **TH+SS**, especially for mmWave (high-band, **HB**) which can deliver ultra-high bandwidth. For example, using S20U, Verizon's mmWave 5G service can provide up to 3 Gbps (see in §3.2). S10 achieves around 2 Gbps over Verizon mmWave 5G (similar to PX5, see Appendix A.1 for details). This highlights the importance of throughput information for the power model construction, especially for mmWave-based networks. Note that there are performance differences between the models constructed using data from different devices (*e.g.,* between first two models). Not surprisingly, this signifies that different devices have different hardware specs (*e.g.,* chipset lithography) that impact power consumption.

Validation on Real Applications. Finally, we evaluate the accuracy of our power model by running two real-world applications: **(1)** video streaming over YouTube app; **(2)** web browsing over Google Chrome Browser app. For video experiments, we play a video [4] at 2K resolution, in both online mode (over cellular radio) and offline mode (downloaded to SD card). To get the network energy consumption, we subtract the total offline energy which contains energy consumed by decoding and rendering of video from the total energy measured when running online. Similarly for web experiments, we download the whole website to SD card

and open the locally stored homepage (`.html`) file on Chrome to load the website in offline mode and then compare the same when loaded in online mode. We compare the energy consumption estimated by our model with the actual energy consumption measured by Monsoon power monitor. The average relative errors are 3.7% for video streaming and 2.1% for web browsing.

4.6 Software power monitor calibration

Although hardware power monitors such as Monsoon [17] provides highly accurate power readings of mobile devices by directly supplying power to them, it will be extremely inconvenient for users to retrieves such information in daily use. In particular, it requires non-trivial hardware engineering efforts on current COTS smartphones (*e.g.,* remove the non-removable back cover and battery). Android exposes battery status such as current (`/sys/class/power_supply/battery/current_now`) and voltage (`/sys/class/power_supply/battery/voltage_now`) which can be used to measure the device power. Thus, besides different impacting factors for power model construction, we also study the accuracy of battery status (current, voltage) readings and whether it can be calibrated and further used to report the device power. First, we run different activities and collect battery status using the software (API) and hardware (Monsoon) and calculate the average relative errors. The software approach always underestimate the power (Table 9 in Appendix A.5). A higher sampling rate may help provide better estimation, but this will incur higher energy overhead (Table 3). Next, we use DTR to calibrate the software power values. Fig. 15 shows the calibration performance (**SW**) together with our **TH+SS** model results. After calibration, the software-based approach can achieve comparable performance. A higher sampling rate (*e.g.,* 10Hz) can even lead to better performance (*i.e.,* lower `MAPE`). However, we argue that a higher sampling rate will incur higher overhead which is less energy-efficient.

To summarize, we empirically characterize several impacting factors such as signal strength and throughput over power consumption by smartphones using 5G services. We propose an ML-based data-driven approach to construct power models for 5G networks. We demonstrate that our models help increase accuracy in predicting device power consumption. We show that the software power monitor can achieve comparable accuracy after calibration.

Next, we take a closer look at two popular mobile applications, video streaming and web browsing, both combined are expected to cover more than 80% of the mobile traffic share by 2025 [8]. We look at them from the perspective of both application QoE and energy efficiency. We believe the proposed power models can be useful to aid developers in making their application more energy-efficient. For the following sections, we focus on mmWave 5G which are considered key to mainstream 5G and have not been studied before in the context of mobile applications.

5 VIDEO STREAMING OVER 5G

Adaptive bitrate (ABR) algorithms are the primary tools used to optimize video quality of experience (QoE). The research community

has proposed a plethora of ABR algorithms for video streaming in recent years [33, 38, 61, 62]. In this section, we demystify 5G's implication on video streaming by conducting the first in-depth investigation of ABR streaming QoE over 5G with mmWave. We aim to answer the following questions:

• What's the performance footprint of the current state-of-the-art ABR algorithms under 5G and how does it compare with 4G?

• What are the major factors that impact ABR streaming performance over 5G?

• What new mechanisms are needed to make future ABR algorithms *5G-aware* and further improve the QoE?

5.1 Evaluation Methodology

Our testbed consists of an Apache server hosting the videos and a DASH.js [15] video client. We use trace-driven emulation to ensure that all algorithms experience the same set of network conditions. We use the Lumos5G dataset [40] which contains 121 5G and 175 4G throughput traces, collected at 1-second granularity. We focus on traces collected with mmWave coverage as 5G's high-band frequency range is considered key to support UHD and beyond video streaming [49]. We use a custom 4K video [3] and encode it using FFmpeg [6] with `libx264` into 6 tracks (or qualities) with different bitrates. 4K (or even 16K) video streaming usually requires 25-120 Mbps (246-328 Mbps) bandwidth [11, 18, 19, 66] which can be easily met by 5G. Thus, to identify rate adaptation challenges in 5G which has a mean throughput value that is 10× of 4G, we scale the video bitrate of 5G tracks used to match its throughput range. This ensures avoiding any trivial bitrate selection. We set the bitrate of the top track (*i.e.,* highest video quality) to match the median throughput of 5G/4G network traces. In this study, the maximum bitrate track for 5G is 160 Mbps, and 20 Mbps for 4G. We then decide the bitrates for lower-quality tracks by keeping the encoded bitrate ratio as ∼1.5 [45] between two adjacent tracks. Note, our goal here is not to understand whether video streaming is better over 5G or 4G. Rather, we focus on studying whether existing ABR algorithms can work well over mmWave 5G. Using the throughput traces, we use Linux `tc` on the client side and control the instantaneous bandwidth. For showing results, we normalize the video bitrates by the bitrate of the top track.

We study the following 7 state-of-the-art ABR algorithms covering 4 different categories. **(1)** *Buffer-based:* BBA [32] and BOLA [56] make bitrate decisions based on the buffer occupancy. **(2)** *Throughput-based:* simple rate-based (RB) and FESTIVE [33] use information of past chunks to estimate future throughput and decide the bitrate of the next chunk to download. **(3)** *Control theoretic:* FastMPC and RobustMPC [62] make bitrate decisions by solving an optimization problem of the QoE for the next *n* chunks (*e.g., n* = 5). **(4)** *Machine learning-based:* Pensieve [38] adopts a deep neural-network to learn bitrate decisions that maximize a QoE reward[7].

5.2 Performance of Existing ABR Schemes

Overall, we find that multiple ABR algorithms that work well under 4G do not maintain the high performance under 5G. Fig. 17 summarizes the bitrate and the video stall time for different ABR

---

[7]We show the results of the Pensieve model trained with real Lumos5G [40] network traces. We also verify that the performance observed by using models trained with synthetic traces (as suggested in their paper [38]) and Lumos5G traces are similar.

algorithms. The top-right rectangular region marked using maroon-colored dashed lines represents ABR algorithms with better QoE. Here, better QoE refers to ABR algorithms that achieve less than 5% video stall and over 0.8 normalized bitrate across different traces. For 5G, only one algorithm (robustMPC) provides better QoE while for 4G there are 3 more algorithms.

Although most of the ABR algorithms under 5G can achieve similar normalized bitrates as they are in 4G (*i.e.,* similar Y-axis values in Figs. 17a and 17b, with an average drop of only 3.5%), the concerning problem for video streaming over 5G lies in the video stalls. For RB, BOLA, MPC, and Pensieve, we observe a significant increase (58.2% on average) of video stall. Fig. 17c shows that except for BBA all other ABR algorithms suffer an increase in video stalls when running over 5G. For instance, the mean video stall time for fastMPC and Pensieve has increased by 82.0% and 259.5%, respectively. Pensieve outperforms all other algorithms in 4G but incurs the highest video stall time under 5G setting. Since Pensieve makes bitrate selection to optimize its QoE reward, we also compare its QoE reward with that of fastMPC and robustMPC. Pensieve's QoE reward improvement is also marginal compared to other algorithms (0.66% improvement over fastMPC and 5.93% over robustMPC), which is 3× lower than the results in the original Pensieve paper. A possible explanation is that for 5G networks, a larger dataset is needed for training the model to learn 5G specific characteristics and make better decisions, which deserves further study. After taking a closer look at the bitrate decisions taken by Pensieve and fastMPC, we find that they sometimes choose the highest bitrate chunk only to regret that it was a wrong decision that is difficult to undo, resulting in a very high stall time. This is not happening in 4G scenarios with the same optimization metric used. Based on this phenomenon, next we dig further to quantitatively understand the challenges involved in running ABR algorithms for video streaming over 5G networks (§5.3).

5.3 Challenges in ABR Streaming under 5G

Throughput prediction. Many ABR algorithms incorporate network throughput into its decision by leveraging a throughput predictor and their performance heavily depends on prediction accuracy. To study the impact of throughput prediction on 5G video streaming, we fix other parts in an ABR algorithm and plug in different throughput predictors and compare the incurred QoE. Considered as one of the state-of-the-art ABR algorithms, we choose fastMPC as the baseline since it explicitly incorporates a throughput predictor while Fugu [61] and Pensieve use throughput information implicitly. We compare three different throughput predictors: **(1)** *hmMPC*: the original throughput predictor used by fastMPC uses harmonic mean of past throughput values to predict future throughput, **(2)** *MPC_GDBT*: a state-of-the-art mmWave 5G-specific throughput predictor [40] that adopts a ML based approach called Gradient Boosted Decision Tree (GDBT), and **(3)** *truthMPC*: ground-truth throughput trace to represent the optimal online throughput prediction scheme. Since MPC's goal is to maximize its QoE function [62], we use the QoE function as the metric to evaluate the effectiveness of applying the 5G-specific throughput predictor. Fig. 18a indicates that using the GDBT throughput predictor can achieve 31.98% higher normalized QoE compared to the default harmonic mean predictor. Compared to *truthMPC* though, adopting the

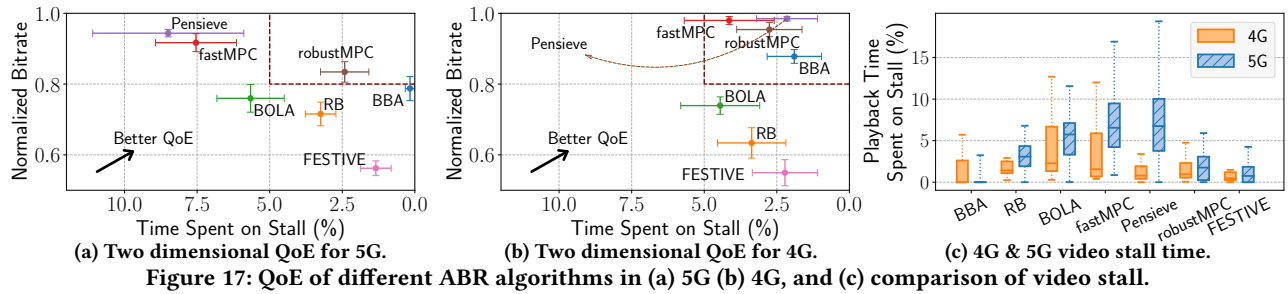(a) Two dimensional QoE for 5G. (b) Two dimensional QoE for 4G. (c) 4G & 5G video stall time.

Figure 17: QoE of different ABR algorithms in (a) 5G (b) 4G, and (c) comparison of video stall.

(a) Throughput Predictor (b) Chunk Length (c) Interface Selection Scheme
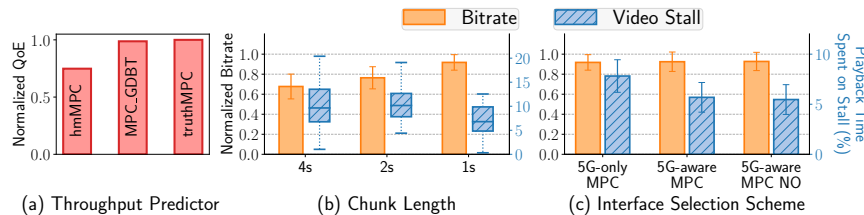
Figure 18: QoE impact of: (a) throughput predictors (b) chunk length, and (c) interface selection schemes.

Table 4: Energy consumption for different interface selection schemes.

Interface selection scheme	Energy (J)
5G-only MPC	495.0±55.1
5G-aware MPC	474.4±59.1
5G-aware MPC NO*	475.0±58.9

(*NO = No Overhead)

Table 5: Factors considered for analyzing their impact on page load time and energy consumption.

Factor	Abbr	Factor	Abbr
# of dynamic/total objs	DNO	# of images (videos)	NI (NV)
Size of dynamic objs / total page size (in bytes)	DSO	Total Page Size	PS
# of objects	NO	Avg. Object Size	AOS

GDBT predictor only provides 1.3% less QoE. Therefore, improving throughput prediction accuracy in ABR algorithms can significantly enhance video streaming QoE and provide opportunities to build better 5G-aware throughput-predictors. Since 5G now spans across many different bands and its network performance variation is large (§3), building better throughput prediction schemes is not only vital to make ABRs work well over 5G but also to improve our understanding of the 5G ecosystem in general.

Decision making granularity. An ABR algorithm's decisions are coarse-grained in that it has to do chunk selection on chunk boundaries, and once made, such decisions cannot be rolled back. Specifically in our 5G video streaming results, we find that just one or two bad chunk selections can significantly affect QoE of the entire stream. This one chunk download decision indeed quickly drains the playback buffer or even causes 5–10 seconds of rebuffering. One fix is to reduce the video chunk length to support fine-grained selections. We study the effect of different chunk lengths (1/2/4s) on 5G video streaming (with fastMPC). Fig. 18b shows that using 1s chunks provides 21.5% (35.9%) higher bitrate and 33.6% (29.8%) less video stalls compared to 2s (4s) chunks. Therefore, although 2s and 4s chunks are typically suggested for ABR [2], we argue video content providers should consider shorter length chunks (*e.g.,* 1s) so that ABR algorithms can make finer-grained decisions and adapt better to the highly fluctuating 5G network conditions.

5.4 Improving 5G ABR Streaming

Based on our observation that 5G consumes more power than 4G when the throughput is low (§4) and 5G throughput fluctuates a lot, we propose **5G-aware video streaming**. The idea is switch to 4G when ABR algorithms predict that 5G throughput is low (*i.e.,* <4G's average throughput), given that 4G provides relatively stable bandwidth, and switch back to 5G when the video buffer level has reached over some threshold (empirically set to 10s). We also take into account the switching overhead between 4G and 5G (§4) and emulate the switching delay using Linux tc. Similarly, we use fastMPC as the baseline ABR algorithm. Figure 18c depicts that our selection scheme (denoted as *5G-aware MPC*) can reduce

video stall time by 26.9% compared to always using 5G interface during the entire video. Compared to the 5G-aware MPC with no overhead version (where we remove the interface switch delay, assuming the UE can instantly switch between 4G and 5G), our realistic interface selection model only incurs 4.0% more stall time. Table 4 shows the corresponding energy consumption, measured by feeding the collected video packet traces into our 5G power model (§4). As shown, the proposed 5G-aware schemes consumes 4.2% less energy than always using 5G. It's also slightly *"greener"* than the no overhead version by trading a little bit of video quality: downloading higher quality chunks and consuming more energy. Figure 18c and Table 4 conclude that carefully selecting between 4G and 5G interfaces can both improve adaptive video streaming performance (26.9% fewer stalls) as well as reduce the energy consumption (by 4.2%, comparative to the 4.7% saving achieved in [59]).

6 QoE IMPLICATIONS OF WEB BROWSING OVER mmWAVE 5G

Previous sections have shown that mmWave 5G is able to provide ultra-high throughput but requires more power to deliver this performance. On the other hand, low-band 5G or LTE uses much less power but delivers lower performance than mmWave. Hence, there is a trade-off between achieving high performance and energy efficiency. To get better insights about this trade-off, in this section we use web browsing as a case study to understand the QoE implications of radio type (*e.g.,* 4G or mmWave 5G) used to load websites in-the-wild.

Data Collection Methodology. Using chrome-har-capturer [14], we build scripts to instrument and load Alexa's top 1500 websites via the Chrome Browser app. For each website, we collect HTTP Archive (*i.e.,* HAR [1]) files as well as capture the packet traces. Since packet capturing requires root permission, we used PX5.We

conduct this experiment under stationary conditions in two radio settings: **(i)** mmWave 5G is active, **(ii)** 4G/LTE is active. mmWave-based experiments were conducted with UE having LoS to 5G tower. We repeat the experiment at least 8 times per device per radio type. To eliminate the impact of browser cache, we clear the cache before loading the next website.

The HAR file of each website loading provides us the total page load time (PLT), time to fetch each individual object (*e.g.,* images, .css or .js files) associated with the website, *etc.* We also extract the per-second throughput trace observed in the packet dumps. This trace is then fed to our power model proposed in §4 to estimate the radio's energy consumption for loading the website. All references to 5G in this section refer to Verizon mmWave 5G service.
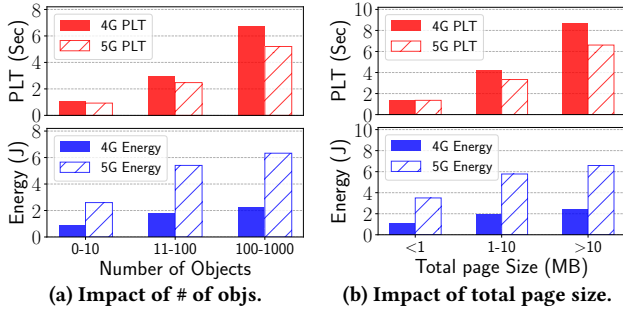
(a) Impact of # of objs.   **(b) Impact of total page size.**

Figure 19: Understanding how different factors affect the page load times under mmWave 5G or 4G setting.

6.1 When does mmWave 5G help?

We list several factors (see Table 5 for the entire list) that might potentially impact PLT performance and/or energy utilization. For each radio type, Fig. 19 compares their empirical impact for a subset of these factors on the two QoE metrics – performance and energy consumption, and makes the following key observations: **(i)** As the number of objects contained in a website increases, the PLT performance gap between 4G and 5G increases with 4G being on the poor side. Similar observations are made for other factors such as total page size, number of dynamic objects. **(ii)** On the other hand, the implications of the very same factors have an opposite effect when seen under the purview of energy consumption where 4G consumes far less energy than 5G. CDF plots in Fig. 20 show these differences more clearly. We find that due to the high-throughput offered by mmWave 5G, PLT performance in 5G is always better than 4G. However, as demonstrated earlier in §4, when applications are not bandwidth-hungry (*e.g.,* normal web browsing), energy utilization of 4G is better than that of mmWave 5G.
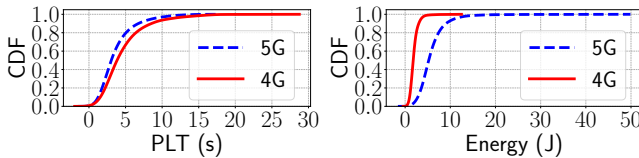
Figure 20: CDF for PLT and energy.

While the importance of performance and energy utilization can differ based on the usage context, we normalize both metrics for fair comparison. Fig. 21 shows that even a 10% penalty over PLT incurred for choosing 4G over 5G can reduce energy consumption by almost 70%. While such a high level of savings diminish as the PLT penalty grows, the important takeaway here is that the slightest

permissible penalty in PLT (caused by choosing 4G) leads to high energy savings. To understand where such a permissible penalty might lie depends on how much additional delay in PLT is permissible such that there is no significant impact on user experiences. For example, a 2s or less PLT remains a widely considered golden standard [44] for web page load times. An average 4G throughput of say 60 Mbps can theoretically load a website with a total page size of 15 *MB* in <2s and might potentially save energy while not significantly affecting the QoE and/or bounce rate.[8]
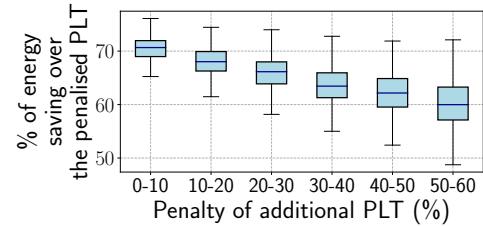
Figure 21: 4G's PLT penalty and energy saving over 5G.

6.2 Interface Selection for Web Browsing

Using all the above insights, we next propose a simple yet insightful model generation algorithm that takes into account all the factors listed in Table 5 to decide whether to use the 4G or 5G radio interface for loading a website. To help make this decision, we come up with a simple linear utility function: QoE = $(\alpha \times EC) + (\beta \times PLT)$ that allows us to tune the weights α and β for the two competing QoE metrics - energy consumption (EC) and page load times (PLT), respectively. To make the generated model insightful, it will be useful to know what factors (from Table 5) of a website makes a model choose a particular radio interface over another.

For this case study, we choose *Decision Tree* (DT) learning algorithm for two reasons. First, DT is easy to run as it does not require any massive computational power. Secondly, it provides indices (*e.g.,* Gini index) for each of the features included in the input feature vector making it easily interpretable. Both these benefits can potentially help application/service developers to not only get insights on improving and achieving their designed QoE but also enable them to account for the usage context and quickly build more models for achieving different QoE goals.

Table 6: DT's radio interface selection results.

#ID	Desired QoE	α	β	Use 4G	Use 5G
M1	High Performance	0.2	0.8	19	401
M2	Performance Oriented	0.4	0.6	366	54
M3	Balanced	0.5	0.5	387	33
M4	Better Energy Saving	0.6	0.4	405	15
M5	High Energy Saving	0.8	0.2	420	0

Model Setup. We randomly split our dataset using a ratio of 7:3 such that 70% is used for training and validation and the rest is used for testing. With over 30K data points, the time to generate the model was less than a minute on a general-purpose laptop.

Results. Table 6 shows the results of different models' radio interface selection results over the 420 websites in the test set. Fig. 22 shows the bottom-up post-pruned DT for models M1 and M4. When performance matters (M1), we find that two factors are important

---

[8]Bounce rate [44]: the percentage of visitors that leave a page without taking an action.
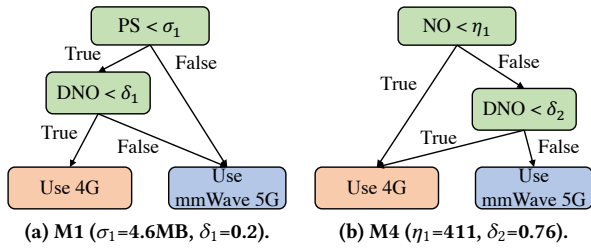
Arvind Narayanan*, Xumiao Zhang*, Ruiyang Zhu, Ahmad Hassan, Shuowei Jin, *et al.*

(a) M1 (σ_1=4.6MB, δ_1=0.2).   **(b) M4 (η_1=411, δ_2=0.76).**

Figure 22: High-Perf (M1) vs. Energy Saving (M4) models.

in deciding the radio type: **(1)** the total page size in bytes, and **(2)** the proportion of dynamic vs. static objects (*e.g.,* ads vs. logos) In contrast, when energy utilization is preferred (M4), 4G radio can handle more websites while 5G will be the preferred radio when the website has an extremely high number of dynamic objects (>76%) compared to static objects. By feeding the web packet traces into our constructed power model (§4), we find that interface selection help save 15-66% energy while improving the overall QoE. The dynamic 4G/5G switching scheme proposed in [59] brings a 25% saving on energy but does not consider the page load time.

7 RELATED WORK

5G Measurements. Xu *et al.* [59] did a measurement study of a commercial mid-band 5G service in China. Narayanan *et al.* [39] established baseline performance of the very initial 5G commercial deployments (mmWave and mid-band) in the US. Lumos5G [40] focused on mmWave 5G throughput characterization and proposed machine learning models for throughput prediction. In our study, we consider both mmWave and low-band 5G with wider-range of 5G smartphone models and server locations. We also conduct the *first* measurement study of an operational SA 5G service.

5G RRC Parameters. Existing work have made various efforts to investigate RRC state machine for 3G/4G [31, 48, 51]. For 5G, Xu *et al.* [59] leverages the UE's diagnostic interface to access lower-layer signaling messages and monitor RRC state transitions. Access to the diagnostic interface requires special license from the chipset vendor which can be challenging and cost-prohibitive. We therefore use an unrooted approach (*i.e.,* RRC-Probe) to infer RRC state machine for both NSA and SA 5G.

5G Power Characteristics. 3G/4G power characteristics have been extensively studied in literature [24, 31, 47, 52] while 5G power characteristics remain under explored. Xu *et al.* [59] conducted a preliminary measurement study to understand mid-band 5G's power consumption and energy efficiency by saturating the link capacity and compare it with that of 4G/LTE. They used a software power monitor to measure power consumption. In this paper, along with a software-based approach, we also use a hardware power monitor to measure power. We provide a more thorough characterization of 5G power consumption for both: mmWave/high-band and low-band 5G, and compare with that of existing 4G/LTE. Our methodology to characterize power consumption includes both: conducting controlled (*e.g.,* at different uplink/downlink target throughput) and in-the-wild (*e.g.,* stationary and walking) experiments.

Smartphone Power Modeling. Prior studies have built power models for 3G [47, 48, 55, 65] and 4G/LTE [26, 31, 42]. Some focus on energy consumption for video streaming [63, 64] and web page load [23]. However, when modeling power during data transfer, they

either treat the network power as a constant value or only consider a single impacting factor such as throughput or signal strength during model construction. For 5G, several factors can together make a significant impact on the smartphone power level, and different 5G technologies relying on different radio frequency bands also incur different power consumption. In this paper, we model the data transfer power considering factors including signal strength and throughput and further show that power model characteristics vary across different 5G bands.

Mobile Video Streaming. Video delivery over LTE has been widely investigated [36, 57, 60, 69]. However, video streaming performance over real commercial 5G networks (especially over mmWave) has been largely under explored. Xu *et al.* [59] performed a preliminary study of UHD panoramic video telephony over mid-band 5G. Han *et al.* [28] showcased an example of streaming volumetric/6D video over a mmWave 5G network under line-of-sight condition. There have also been efforts on evaluating the performance of different ABR algorithms for HTTP adaptive streaming [21, 37, 38, 61]. Nevertheless, none of them have examined the performance of existing ABR algorithms over 5G. Researchers also observe that better throughput prediction can improve video performance in cellular networks [71]. This is even more important for 5G ecosystem that supports a wide range of frequency bands with diverse coverage and performance characteristics. For instance, in the case of mmWave 5G, performance can be greatly affected by environmental and user-side factors [39, 40]. Xu *et al.* [59] did a preliminary study on optimizing 5G power management by dynamically switching between 4G and 5G interfaces. However, their goal was to solely improve energy efficiency, not for application QoE.

Mobile Web Browsing. Previous studies mostly focus on understanding and improving web page loading over legacy 3G/4G networks [22, 46, 58]. Narayanan *et al.* [39] studied web page loading performance with different HTTP protocol version numbers and encryption configurations using mmWave 5G. Xu *et al.* [59] looked into the downloading and rendering performance with different types of websites using mid-band 5G. Our study includes a comprehensive examination of the performance and energy consumption loading top websites using 5G, and proposes simple yet intelligent interface selection schemes to satisfy different QoE goals.

8 CONCLUSION

Leveraging a custom measurement platform, we have conducted comprehensive measurements of several key aspects of commercial 5G: end-to-end network performance, power characteristics, 4G/5G interaction, and application QoE. Our findings reveal the state-of-the-art landscape of the 5G ecosystem, in particular the higher protocol stack. We have released our datasets and measurement tools to the research community.

ACKNOWLEDGMENTS

REFERENCES

[1] 2007. HAR 1.2 Spec. (2007). http://www.softwareishard.com/blog/har-12-spec/
[2] 2015. Optimal Adaptive Streaming Formats MPEG-DASH & HLS Segment Length. (2015). Retrieved January 2021 from https://bitmovin.com/mpeg-dash-hls-segment-length/
[3] 2016. Real 4K HDR 60fps: LG Jazz HDR UHD (Chromecast Ultra). (2016). https://www.youtube.com/watch?v=mkggXE5e2yk
[4] 2018. Cobra Kai Ep 1 – "Ace Degenerate" – The Karate Kid Saga Continues. (2018). https://www.youtube.com/watch?v=_rB36UGoP4Y
[5] 2019. ETSI TR 103 559: Speech and multimedia Transmission Quality (STQ); Best practices for robust network QoS benchmark testing and scoring. (2019). https://www.etsi.org/deliver/etsi_tr/103500_103599/103559/01.01.01_60/tr_103559v010101p.pdf
[6] 2019. FFmpeg Project. (2019). Retrieved January 2021 from http://ffmpeg.org/
[7] 2019. Snapdragon X50 5G modem-RF system. (2019). Retrieved January 2021 from https://www.qualcomm.com/products/snapdragon-x50-5g-modem
[8] 2020. Ericsson Mobility Report. (2020). https://www.ericsson.com/49da93/assets/local/mobility-report/documents/2020/june2020-ericsson-mobility-report.pdf
[9] 2020. Snapdragon 765G 5G Mobile Platform. (2020). Retrieved January 2021 from https://www.qualcomm.com/products/snapdragon-765g-5g-mobile-platform
[10] 2020. Snapdragon X55 5G modem-RF system. (2020). Retrieved January 2021 from https://www.qualcomm.com/products/snapdragon-x55-5g-modem
[11] 2020. What's the Best Bitrate for the Best Video Quality on YouTube? (1080p, 1440p, 4K). (2020). https://www.youtube.com/watch?v=0fz479id_Ic
[12] 2021. 3GPP TS 36.331: Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification (V16.3.0). (2021).
[13] 2021. 3GPP TS 38.331: NR; Radio Resource Control (RRC); Protocol specification (V16.3.1). (2021).
[14] 2021. Chrome HAR capturer. (2021). https://github.com/cyrus-and/chrome-har-capturer
[15] 2021. Dash-Industry-Forum, dash.js. (2021). https://github.com/Dash-Industry-Forum/dash.js
[16] 2021. Linux CUBIC TCP. (2021). https://github.com/torvalds/linux/blob/master/net/ipv4/tcp_cubic.c
[17] 2021. Monsoon power monitor. https://www.msoon.com/LabEquipment/PowerMonitor/. (2021).
[18] 2021. Netflix recommended network bandwidth for 4K video. (2021). https://help.netflix.com/en/node/306
[19] 2021. YouTube 4K bitrates encoding. (2021). https://support.google.com/youtube/answer/1722171
[20] 3rd Generation Partnership Project. 2019. Release 15. (April 2019). Retrieved January 2021 from https://www.3gpp.org/release-15
[21] Zahaib Akhtar, Yun Seong Nam, Ramesh Govindan, Sanjay Rao, Jessica Chen, Ethan Katz-Bassett, Bruno Ribeiro, Jibin Zhan, and Hui Zhang. 2018. Oboe: auto-tuning video abr algorithms to network conditions. In *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*. 44–58.
[22] Duc Hoang Bui, Yunxin Liu, Hyosu Kim, Insik Shin, and Feng Zhao. 2015. Rethinking energy-performance trade-off in mobile web page loading. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. 14–26.
[23] Yi Cao, Javad Nejati, Muhammad Wajahat, Aruna Balasubramanian, and Anshul Gandhi. 2017. Deconstructing the energy consumption of the mobile page load. *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 1, 1 (2017), 1–25.
[24] Xiaomeng Chen, Ning Ding, Abhilash Jindal, Y Charlie Hu, Maruti Gupta, and Rath Vannithamby. 2015. Smartphone energy drain in the wild: Analysis and implications. *ACM SIGMETRICS Performance Evaluation Review* 43, 1 (2015), 151–164.
[25] Shuguang Cui, Andrea J Goldsmith, and Ahmad Bahai. 2004. Energy-efficiency of MIMO and cooperative MIMO techniques in sensor networks. *IEEE Journal on selected areas in communications* 22, 6 (2004), 1089–1098.
[26] Ning Ding, Daniel Wagner, Xiaomeng Chen, Abhinav Pathak, Y Charlie Hu, and Andrew Rice. 2013. Characterizing and modeling the impact of wireless signal strength on smartphone battery drain. *ACM SIGMETRICS Performance Evaluation Review* 41, 1 (2013), 29–40.
[27] Signals Research Group. 2019. A Global Perspective of 5G Network Performance. (2019). Retrieved June 2020 from https://www.qualcomm.com/media/documents/files/signals-research-group-s-5g-benchmark-study.pdf
[28] Bo Han, Yu Liu, and Feng Qian. 2020. ViVo: Visibility-aware mobile volumetric video streaming. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–13.
[29] K. Heimann, J. Tiemann, D. Yolchyan, and C. Wietfeld. 2019. Experimental 5G mmWave Beam Tracking Testbed for Evaluation of Vehicular Communications. In *2019 IEEE 2nd 5G World Forum (5GWF)*. 382–387. https://doi.org/10.1109/5GWF.2019.8911692
[30] Anders Hillbur. 2018. 5G deployment options to reduce the complexity. (2018). Retrieved January 2021 from https://www.ericsson.com/en/blog/2018/11/5g-deployment-options-to-reduce-the-complexity

[31] Junxian Huang, Feng Qian, Alexandre Gerber, Z Morley Mao, Subhabrata Sen, and Oliver Spatscheck. 2012. A close examination of performance and power characteristics of 4G LTE networks. In *Proceedings of the 10th international conference on Mobile systems, applications, and services*. 225–238.
[32] Te-Yuan Huang, Ramesh Johari, Nick McKeown, Matthew Trunnell, and Mark Watson. 2014. A buffer-based approach to rate adaptation: Evidence from a large video streaming service. In *Proceedings of the 2014 ACM conference on SIGCOMM*. 187–198.
[33] Junchen Jiang, Vyas Sekar, and Hui Zhang. 2012. Improving fairness, efficiency, and stability in http-based adaptive video streaming with festive. In *Proceedings of the 8th international conference on Emerging networking experiments and technologies*. 97–108.
[34] Wooseong Kim. 2019. Experimental Demonstration of MmWave Vehicle-to-Vehicle Communications Using IEEE 802.11 ad. *Sensors* 19, 9 (2019), 2057.
[35] K. Larsson, B. Halvarsson, D. Singh, R. Chana, J. Manssour, M. Na, C. Choi, and S. Jo. 2017. High-Speed Beam Tracking Demonstrated Using a 28 GHz 5G Trial System. In *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*. 1–5. https://doi.org/10.1109/VTCFall.2017.8288043
[36] Jinsung Lee, Sungyong Lee, Jongyun Lee, Sandesh Dhawaskar Sathyanarayana, Hyoyoung Lim, Jihoon Lee, Xiaoqing Zhu, Sangeeta Ramakrishnan, Dirk Grunwald, Kyunghan Lee, et al. 2020. PERCEIVE: deep learning-based cellular uplink prediction using real-time scheduling patterns. In *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services*. 377–390.
[37] Melissa Licciardello, Maximilian Grüner, and Ankit Singla. 2020. Understanding video streaming algorithms in the wild. In *International Conference on Passive and Active Network Measurement*. Springer, 298–313.
[38] Hongzi Mao, Ravi Netravali, and Mohammad Alizadeh. 2017. Neural adaptive video streaming with pensieve. In *Proceedings of the Conference of the ACM Special Interest Group on Data Communication*. 197–210.
[39] 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 *Proceedings of The Web Conference 2020 (WWW '20)*. Association for Computing Machinery, New York, NY, USA, 894–905. https://doi.org/10.1145/3366423.3380169
[40] Arvind Narayanan, Eman Ramadan, Rishabh Mehta, Xinyue Hu, Qingxu Liu, Rostand AK Fezeu, Udhaya Kumar Dayalan, Saurabh Verma, Peiqi Ji, Tao Li, et al. 2020. Lumos5G: Mapping and Predicting Commercial mmWave 5G Throughput. In *Proceedings of the ACM Internet Measurement Conference*. 176–193.
[41] Arvind Narayanan, Eman Ramadan, Jacob Quant, Peiqi Ji, Feng Qian, and Zhi-Li Zhang. 2020. 5G Tracker – A Crowdsourced Platform to Enable Research Using Commercial 5G Services. In *Proceedings of the ACM SIGCOMM 2020 Conference Posters and Demos (SIGCOMM Posters and Demos '20)*. Association for Computing Machinery, Virtual Event, USA. https://doi.org/10.1145/3405837.3411394
[42] Ana Nika, Yibo Zhu, Ning Ding, Abhilash Jindal, Y Charlie Hu, Xia Zhou, Ben Y Zhao, and Haitao Zheng. 2015. Energy and performance of smartphone radio bundling in outdoor environments. In *Proceedings of the 24th International Conference on World Wide Web*. 809–819.
[43] OOKLA. [n. d.]. Speedtest® for Android. ([n. d.]). Retrieved January 2021 from https://www.speedtest.net/apps/android
[44] Pingdom. 2018. Does Page Load Time Really Affect Bounce Rate? | Pingdom. (2018). Retrieved January 2021 from https://www.pingdom.com/blog/page-load-time-really-affect-bounce-rate/
[45] Feng Qian, Bo Han, Qingyang Xiao, and Vijay Gopalakrishnan. 2018. Flare: Practical viewport-adaptive 360-degree video streaming for mobile devices. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*. 99–114.
[46] Feng Qian, Subhabrata Sen, and Oliver Spatscheck. 2014. Characterizing resource usage for mobile web browsing. In *Proceedings of the 12th annual international conference on Mobile systems, applications, and services*. 218–231.
[47] Feng Qian, Zhaoguang Wang, Alexandre Gerber, Zhuoqing Mao, Subhabrata Sen, and Oliver Spatscheck. 2011. Profiling resource usage for mobile applications: a cross-layer approach. In *Proceedings of the 9th international conference on Mobile systems, applications, and services*. 321–334.
[48] Feng Qian, Zhaoguang Wang, Alexandre Gerber, Zhuoqing Morley Mao, Subhabrata Sen, and Oliver Spatscheck. 2010. Characterizing radio resource allocation for 3G networks. In *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*. 137–150.
[49] Eman Ramadan, Arvind Narayanan, Udhaya K. Dayalan, Rostand A. K. Fezeu, Feng Qian, and Zhi-Li Zhang. 2021. Case for 5G-Aware Video Streaming Applications. *ACM SIGCOMM'21 5G-MeMU Workshop* (2021).
[50] Theodore S Rappaport, George R MacCartney, Mathew K Samimi, and Shu Sun. 2015. Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design. *IEEE transactions on Communications* 63, 9 (2015), 3029–3056.
[51] Sanae Rosen, Haokun Luo, Qi Alfred Chen, Z Morley Mao, Jie Hui, Aaron Drake, and Kevin Lau. 2014. Discovering fine-grained RRC state dynamics and performance impacts in cellular networks. In *Proceedings of the 20th annual international*

Arvind Narayanan*, Xumiao Zhang*, Ruiyang Zhu, Ahmad Hassan, Shuowei Jin, *et al.*

conference on Mobile computing and networking. 177–188.

[52] Sanae Rosen, Ashkan Nikravesh, Yihua Guo, Z Morley Mao, Feng Qian, and Subhabrata Sen. 2015. Revisiting network energy efficiency of mobile apps: Performance in the wild. In *Proceedings of the 2015 Internet Measurement Conference.* 339–345.

[53] J. Sachs, G. Wikstrom, T. Dudda, R. Baldemair, and K. Kittichokechai. 2018. 5G Radio Network Design for Ultra-Reliable Low-Latency Communication. *IEEE Network* 32, 2 (2018), 24–31. https://doi.org/10.1109/MNET.2018.1700232

[54] Samsung. 2017. 4G-5G Interworking. (2017). Retrieved January 2021 from https://images.samsung.com/is/content/samsung/p5/global/business/networks/insights/white-paper/4g-5g-interworking/global-networks-insight-4g-5g-interworking-0.pdf

[55] Aaron Schulman, Vishnu Navda, Ramachandran Ramjee, Neil Spring, Pralhad Deshpande, Calvin Grunewald, Kamal Jain, and Venkata N Padmanabhan. 2010. Bartendr: a practical approach to energy-aware cellular data scheduling. In *Proceedings of the sixteenth annual international conference on Mobile computing and networking.* 85–96.

[56] Kevin Spiteri, Rahul Urgaonkar, and Ramesh K Sitaraman. 2016. BOLA: Near-optimal bitrate adaptation for online videos. In *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications.* IEEE, 1–9.

[57] Zhaowei Tan, Yuanjie Li, Qianru Li, Zhehui Zhang, Zhehan Li, and Songwu Lu. 2018. Supporting mobile VR in LTE networks: How close are we? *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 2, 1 (2018), 1–31.

[58] Xiufeng Xie, Xinyu Zhang, and Shilin Zhu. 2017. Accelerating mobile web loading using cellular link information. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services.* 427–439.

[59] Dongzhu Xu, Anfu Zhou, Xinyu Zhang, Guixian Wang, Xi Liu, Congkai An, Yiming Shi, Liang Liu, and Huadong Ma. 2020. Understanding Operational 5G: A First Measurement Study on Its Coverage, Performance and Energy Consumption. In *Proceedings of the Annual conference of the ACM Special Interest Group on Data Communication on the applications, technologies, architectures, and protocols for computer communication.* 479–494.

[60] Shichang Xu, Subhabrata Sen, Z Morley Mao, and Yunhan Jia. 2017. Dissecting VOD services for cellular: performance, root causes and best practices. In *Proceedings of the 2017 Internet Measurement Conference.* 220–234.

[61] Francis Y Yan, Hudson Ayers, Chenzhi Zhu, Sadjad Fouladi, James Hong, Keyi Zhang, Philip Levis, and Keith Winstein. 2020. Learning in situ: a randomized experiment in video streaming. In *17th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 20).* 495–511.

[62] Xiaoqi Yin, Abhishek Jindal, Vyas Sekar, and Bruno Sinopoli. 2015. A control-theoretic approach for dynamic adaptive video streaming over HTTP. In *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication.* 325–338.

[63] Chaoqun Yue, Subhabrata Sen, Bing Wang, Yanyuan Qin, and Feng Qian. 2020. Energy considerations for ABR video streaming to smartphones: Measurements, models and insights. In *Proceedings of the 11th ACM Multimedia Systems Conference.* 153–165.

[64] Jingyu Zhang, Gan Fang, Chunyi Peng, Minyi Guo, Sheng Wei, and Viswanathan Swaminathan. 2016. Profiling energy consumption of DASH video streaming over 4G LTE networks. In *Proceedings of the 8th International Workshop on Mobile Video.* 1–6.

[65] Lide Zhang, Birjodh Tiwana, Zhiyun Qian, Zhaoguang Wang, Robert P Dick, Zhuoqing Morley Mao, and Lei Yang. 2010. Accurate online power estimation and automatic battery behavior based power model generation for smartphones. In *Proceedings of the eighth IEEE/ACM/IFIP international conference on Hardware/software codesign and system synthesis.* 105–114.

[66] Wenxiao Zhang, Feng Qian, Bo Han, and Pan Hui. 2021. DeepVista: 16K Panoramic Cinema on Your Mobile Device. In *Proceedings of the Web Conference 2021 (WWW '21).* Association for Computing Machinery, New York, NY, USA, 2232–2244. https://doi.org/10.1145/3442381.3449829

[67] Hang Zhao, Rimma Mayzus, Shu Sun, Mathew Samimi, Jocelyn K Schulz, Yaniv Azar, Kevin Wang, George N Wong, Felix Gutierrez Jr, and Theodore S Rappaport. 2013. 28 GHz millimeter wave cellular communication measurements for reflection and penetration loss in and around buildings in New York city.. In *ICC.* 5163–5167.

[68] K. Zhao, J. Helander, D. Sjöberg, S. He, T. Bolin, and Z. Ying. 2017. User Body Effect on Phased Array in User Equipment for the 5G mmWave Communication System. *IEEE Antennas and Wireless Propagation Letters* 16 (2017), 864–867.

[69] Xiao Zhu, Subhabrata Sen, and Z Morley Mao. 2021. Livelyzer: analyzing the first-Mile ingest performance of live video streaming. In *Proceedings of the 12th ACM Multimedia Systems Conference.*

[70] Yibo Zhu, Zengbin Zhang, Zhinus Marzi, Chris Nelson, Upamanyu Madhow, Ben Y Zhao, and Haitao Zheng. 2014. Demystifying 60GHz outdoor picocells. In *Proceedings of the 20th annual international conference on Mobile computing and networking.* 5–16.

[71] Xuan Kelvin Zou, Jeffrey Erman, Vijay Gopalakrishnan, Emir Halepovic, Rittwik Jana, Xin Jin, Jennifer Rexford, and Rakesh K Sinha. 2015. Can accurate predictions improve video streaming in cellular networks?. In *Proceedings of the 16th*

International Workshop on Mobile Computing Systems and Applications. 57–62.

A APPENDICES

Appendices are supporting material that has not been peer-reviewed.

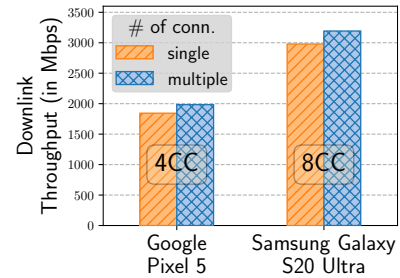## A.1 Impact of UE-Specs and Capabilities

Figure 23: Support for improved carrier aggregation schemes in 5G-NR radios boost throughput performance.

Commercial 5G landscape has improved over time along several dimensions. Most notably for this experiment that tries to quantify the impact of UE-specs on network performance, we find that latest high-end smartphones such as S20U are able to improve downlink and uplink throughput by increasing the number of radio channels (often referred to as carrier aggregation) used between the UE and RAN. For example, previous generation of 5G smartphones (*e.g.,* considered in the baseline w/ QC X50 modem [7]) as well as the cheaper variants of mmWave 5G phones (*e.g.,* PX5 w/ QC X52 modem [9]) uses 4×100 MHz or 4CC (component carriers) for downlink data transfers and 1CC for uplink. On the other hand, S20U (w/ QC X55 modem [10]) supports 8CC over downlink (and 2CC over uplink) resulting in significant improvements in throughput performance. Fig. 23 compares the downlink and uplink throughput between PX5 and S20U. Clearly, S20U provides 50% to 60% improvements in both uplink and downlink throughput over PX5 and the baseline. Of course, harnessing for such carrier aggregation schemes over mmWave bands also requires support from 5G carriers and their infrastructure. We did not find any significant impact of UE specs over latency.

A.2 Impact of Server-Side & Other Factors

Due to mmWave 5G's ultra-high throughput, performance bottleneck can also be due to factors at end-devices. While we just illustrated this on the UE-side, we now take a look at how server-side and other factors might affect mmWave 5G's performance. We have already seen earlier (§3.2) that the default Linux kernel's TCP parameters such as `tcp_wmem` on the server-side need to be increased drastically to improve single TCP connection throughput and utilize mmWave 5G's available bandwidth. Similarly, in the future, when uplink throughput improves and more generically for any transport protocol, careful attention is needed to ensure kernel parameters on both (server and UE) ends are tuned to support application needs and fully utilize mmWave's ultra-high bandwidth capacity. However, challenges lie to ensure such changes to transport layer configurations do not adversely affect other connections over the network that might not necessarily support or need such high throughput.

Table 7: Important 4G/5G RRC parameters using RRC-Probe.

| Mobile Service		RRC Parameter (ms)				
Carrier	Radio type	UE-inactivity timer	Long DRX cycle	IDLE DRX cycle	4G promotion delay	5G promotion delay
T-Mobile	SA low-band	10400	40	1250	N/A	341
T-Mobile	NSA low-band	10400 (12120)	320	1200	210	1440
Verizon	NSA mmWave	10500	320	1280	396	1907
Verizon	NSA low-band	10200 (18800)	400	1100	288	N/A
T-Mobile	4G	5000	400	1300	190	N/A
Verizon	4G	10200	300	1280	265	N/A
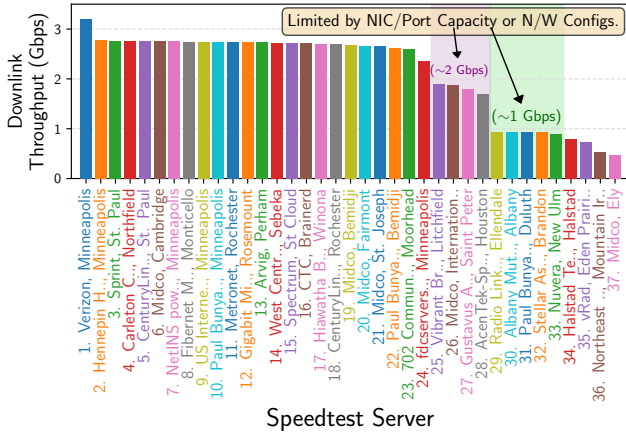
Figure 24: [Verizon mmWave] UE's downlink throughput perceived using several Speedtest servers located in the same state as that of UE (Minnesota). Using Verizon's own server located in the UE's local city (Minneapolis) achieves highest throughput. Others are affected by other (*e.g.,* Internet-side or server-side) factors, *e.g.,* NIC/Switch-Port capacity, network configurations and/or congestion.

Next, we try to understand how different Speedtest servers located in the same state as that of UE (*i.e.,* Minnesota) impact throughput performance. Such bandwidth testing servers are typically hosted by ISPs, mobile operators, and academic organizations. Fig. 24 shows the UE's downlink throughput (using multiple connections) for all the servers. No doubt, the carrier's own hosted server (Verizon) provides the best throughput of over 3 Gbps. Servers 2 to 23 also provide an impressive downlink throughput of ~2.8 Gbps (*i.e.,* 10% degradation over Verizon's own server). This is most likely due to the additional Internet side routing overhead which also increases latency. We also find evidence that production level Speedtest servers might actually not support throughput over certain limits. For instance, we find servers 25 to 28 are bound by 2 Gbps, while servers 29 to 33 are bound by 1 Gbps. We believe these bounds might either be due to NIC/switch-port limitations or network configurations. In either case, with mmWave's ultra-high throughput capacity, servers should also have sufficient uplink/downlink capacities to the Internet which can be challenging due to increase in costs and/or infrastructure limitations.

A.3 RRC State Machine Parameters

We summarize a list of timers of RRC state transitions for different networks/carriers/band configurations in Table 7. When the radio is active and there are no incoming/outgoing packets, UE starts the tail timer (*i.e.,* UE-inactivity timer) and stays in RRC_CONNECTED
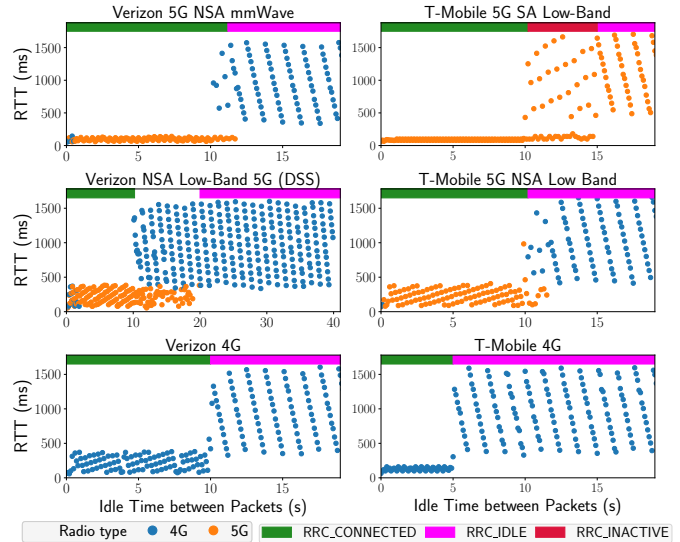

Figure 25: Results of inferring different RRC States using RRC-Probe for SA 5G, NSA 5G and 4G/LTE.

for T_{tail} before demoting to RRC_IDLE. Discontinuous Reception (DRX) is adopted by both 4G and 5G for power saving in which UE periodically wakes up to check paging messages and rests for the remaining time of the cycle. The periods in RRC_CONNECTED and RRC_IDLE are different. T_{long_drx} is the cycle period of Long DRX in RRC_CONNECTED and T_{idle_drx} is the cycle period of DRX in RRC_IDLE. We do not observe and infer Short DRX cycle with RRC-Probe due to its very small cycle period. We also calculate the delay for promotion from RRC_IDLE to 4G and 5G which is T_{4g_pro} and T_{5g_pro} respectively. Fig. 25 shows the results of the different RRC states inferred using RRC-Probe for all the configurations. Note, we observe that in NSA, sometimes the packets might arrive over 4G interface (with higher latency) while other times packets might arrive over 5G interface (with lower latency). This can be seen for the NSA low-band 5G setting for both Verizon and T-Mobile carriers. We have therefore also mentioned a second tail-timer for such settings (see timers in brackets in Table 7). Although not shown, for 4G → 5G promotion in NSA 5G, UE will first promote to 4G's CONNECTED state before switching to 5G (*i.e.,* LTE_RRC_IDLE → LTE_RRC_CONNECTED → NR_RRC_CONNECTED). In SA 5G though, the UE will directly directly reach NR_RRC_CONNECTED.

A.4 Data Transfer (Throughput vs. Power)

4G vs. 5G. Similar to Fig.11 in §4.3, which reports the throughput-power relationship for mmWave 5G, low-band 5G, and 4G using
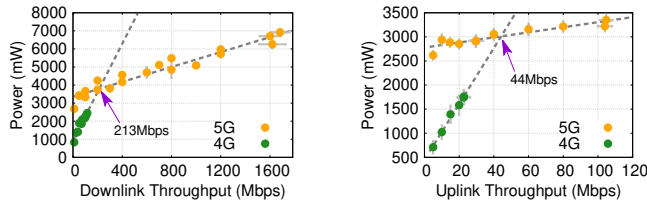
Arvind Narayanan*, Xumiao Zhang*, Ruiyang Zhu, Ahmad Hassan, Shuowei Jin, *et al.*

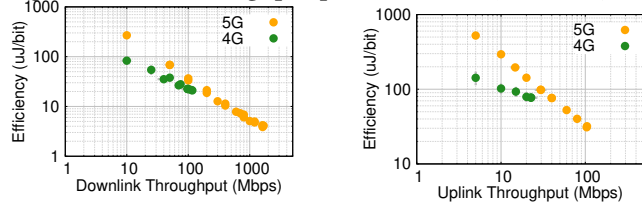**Figure 26: (S10) Throughput-power for 4G & (mmWave) 5G.**

Figure 27: (S10) Throughput-energy for 4G & (mmWave) 5G.

S20U in Minneapolis (MN), we also conduct the same set of experiments using S10 smartphones in Ann Arbor (MI), which have relatively older 5G modems and chipsets. Fig. 26 shows the throughput-power relationship of mmWave 5G and 4G for both downlink and uplink data transfer at controlled throughput target levels. For the downlink and uplink transfer, we echo the observations made earlier in §4.3 that mmWave 5G uses more power than 4G/LTE at low throughput levels, but mmWave becomes more efficient at higher throughput levels. The throughput-energy efficiency results are shown in Fig. 27. Besides, as reported in [24], we also find that the power consumption across different UE models can be different. For example, the crossover points between mmWave 5G and 4G/LTE observed using S10 are different from those measured using S20U. Nonetheless, the crossover points between S10 and S20U are reasonably close to each other.

Table 8: Slopes of Throughput-Power curves indicating increase in power for every 1 Mbps rise in throughput.

Device	Network	Downlink (mW/Mbps)	Uplink (mW/Mbps)
S10	4G	13.38	57.99
S10	5G (mmWave)	2.06	5.27
S20U	4G	14.55	80.21
S20U	5G (low-band)	13.52	29.15
S20U	5G (mmWave)	1.81	9.42

Downlink vs. Uplink. We also compare the downlink transfer with uplink transfer for 4G and 5G. From the results seen in Figs. 11 and 26, we derive the slopes of throughput-power curves across different device models and radio bands/technologies and list them in Table 8 for different settings. From the results, we conclude that uplink power increases 2.2× to 5.9× faster than downlink power for both 5G and 4G, and downlink transfer is always more efficient than uplink. This aligns with previous results on 3G/4G [31]. Unsurprisingly, UE's radio requires more power for sending data than to receive [25]. We have quantitatively compared them between state-of-the-art 5G and 4G commercial services.

A.5 Benchmarking Software-based Power Monitor

We benchmark the software-based power monitor with different activities including **(1)** randomly tapping on the screen and opening/closing applications, **(2)** leaving the UE idle with the screen

Table 9: Benchmarking results on different test cases.

Test Case	Relative error = SW / HW	
	@ 1Hz	@ 10Hz
Random activities	84.2%	94.3%
Idle (screen on)	87.9%	93.7%
Idle (screen off)	80.9%	94.9%
UDP DL 50Mbps	87.1%	91.5%
UDP DL 400Mbps	87.4%	89.7%
UDP DL 800Mbps	87.5%	91.3%
UDP DL 1200Mbps	86.8%	91.2%
Video streaming	92.2%	92.9%

on/off, **(3)** performing UDP download at different speeds, and **(4)** running a video playback. We collect the battery status using both software (API) and hardware (Monsoon) approaches and calculate the average relative errors between the two approaches. The results are shown in Table 9. The software monitor always underestimates the UE power but a higher sampling rate may reduce the error.

A.6 Summary of Artifacts

The GitHub repository mentioned below contains the artifacts (dataset and tools) associated with the paper:

<center>https://github.com/SIGCOMM21-5G/artifact</center>

This is a measurement paper with several types of experiments conducted for different purposes having different methodologies. To help quickly navigate and have the ability to understand the different pieces, we have created different folders for different experiments. There are README files within each folder that provide instructions on validating the experiment-specific artifacts. At the very top of the README instructions, we also specify which results/plots the folder is corresponds to. Lastly, here are some generic principles we followed for releasing the artifacts:

A.6.1 Dataset Size.

(1) If the dataset is small enough, we included the dataset file in the repository itself.
(2) If the dataset files are huge, we use a small sample of the dataset in the repository to demonstrate the functionality.
(3) You can replace the small subset with the full dataset. The full dataset is provided in the experiment-specific README file. In either case, we provide full processed results as well.

A.6.2 Data Analysis, Model/Plot Generation.

(1) If data analysis is involved, our instructions will contain information on how to process the data.
(2) No matter what the dataset size is, we provide the fully generated results and/or plots. If you decide to run the analysis and/or plotting scripts, the outcome of processing will replace the existing files in the repository.
(3) For the artifacts involved in §5 (ABR video streaming), extensive computation resources are required. We have therefore provide a screencast to show how the results were generated. If one can arrange their own compute resources, we provide instructions on how to setup the system and evaluate.

If you have any questions, feel free to reach out to the corresponding authors: arvind@cs.umn.edu, xumiao@umich.edu.