Dissecting Carrier Aggregation in 5G Networks: Measurement, QoE Implications and Prediction

Wei Ye[†], Xinyue Hu[†], Steven Sleder[†], Anlan Zhang[‡], Udhaya Kumar Dayalan[†], Ahmad Hassan[‡], Rostand A. K. Fezeu[†], Akshay Jajoo[*], Myungjin Lee[*], Eman Ramadan[†], Feng Qian[‡], Zhi-Li Zhang[†]
[†]University of Minnesota – Twin Cities [‡]University of Southern California [*]Cisco Research

ABSTRACT

By aggregating multiple channels, Carrier Aggregation (CA) is an important technology for boosting cellular network bandwidth. Given diverse radio bands made available in 5G networks, CA plays a particularly critical role in achieving the goal of multi-Gbps throughput performance. In this paper, we carry out a timely comprehensive measurement study of CA deployment in commercial 5G networks (as well as 4G networks). We identify the key factors that influence whether CA is deployed and when, as well as which band combinations are used. Thus, we reveal the challenges posed by CA in 5G performance analysis and prediction as well as their implications in application quality-of-experience (QoE). We argue for and develop a novel *CA-aware* deep learning framework, dubbed *Prism5G*, which explicitly accounts for the complexity introduced by CA to more effectively predict 5G network throughput performance. Through extensive evaluations, we demonstrate the superiority of *Prism5G* over existing throughput prediction algorithms. *Prism5G* improves 5G throughput prediction accuracy by over 14% on average and a maximum of 22%. Using two use cases as examples, we further illustrate how *Prism5G* can aid applications in optimizing QoE performance.

CCS CONCEPTS

• **Networks** → **Network measurement**; **Network performance analysis**; **Mobile networks**; • **Computing methodologies** → **Machine learning**.

KEYWORDS

Carrier Aggregation, 4G, 5G, Network Measurement, Mobile Network Throughput Prediction, Deep Learning

ACM Reference Format:
Wei Ye[†], Xinyue Hu[†], Steven Sleder[†], Anlan Zhang[‡], Udhaya Kumar Dayalan[†], Ahmad Hassan[‡], Rostand A. K. Fezeu[†], Akshay Jajoo[*], Myungjin Lee[*], Eman Ramadan[†], Feng Qian[‡], Zhi-Li Zhang[†]. 2024. Dissecting Carrier Aggregation in 5G Networks: Measurement, QoE Implications and Prediction. In *ACM SIGCOMM 2024 Conference (ACM SIGCOMM '24), August 4–8, 2024, Sydney, NSW, Australia.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3651890.3672250

Corresponding authors: ye000094@umn.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
ACM SIGCOMM '24, August 4–8, 2024, Sydney, NSW, Australia
© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-0614-1/24/ 08.
https://doi.org/10.1145/3651890.3672250

1 INTRODUCTION

5G was designed to deliver significantly higher data rates than 4G, with a target downlink (DL) peak data rate of 20 Gbps [19]. To achieve this goal, 5G employs a variety of different mechanisms, many of which build on those that have been deployed in 4G networks. First of all, besides the frequency bands in the low- (< 1 GHz) and mid-band (1 GHz – 7 GHz) range (frequency range 1 or FR1) that overlaps with the 4G frequency bands, 5G also utilizes high-band frequencies (24 GHz – 60 GHz) in the mmWave range (frequency range 2 or FR2). Secondly, instead of a fixed 15 KHz sub-carrier-spacing (SCS) and a maximum channel bandwidth of 20 MHz, 5G introduces a flexible *numerology* to increase spectral efficiency, allowing 15/30/60 KHz SCS for FR1 bands and up to 100 MHz channel bandwidth, and 120/240 SCS and up to 400 MHz channel bandwidth for most FR2 bands. Thirdly, massive MIMO (multiple inputs, multiple outputs) may be used to increase the data rate by transmitting up to 4 (or in some cases 8) data streams simultaneously. In particular, carrier aggregation (CA), which combines multiple channels – each is referred to as "component carrier" (CC) – within the same band or across different bands (**§2.1**), plays a crucial role in boosting 5G data rates to multi-Gbps. We use *x CCs* to denote *x* number of channels being aggregated for simplicity.

While CA has been deployed in 4G networks, more diverse band combinations, together with flexible numerology, generally wider channel bandwidths, as well as MIMO enable 5G networks to achieve significantly higher data rates. To illustrate the comparative throughput performance of 4G vs. 5G networks, in Fig. 1, we plot the representative measurement results in the ideal channel condition. We see that the throughput performance of 5G networks is significantly higher than 4G networks. As of Jan 2024, we have observed up to 4 CCs aggregation in 5G low/mid-band and 8 CCs aggregation in 5G high-band, achieving an unprecedented *1.7 Gbps* and *4.1 Gbps* peak throughput performance in the wild, respectively. Moreover, with 5G stand-alone (SA) networks no longer relying on a 4G as an "anchor" cell, CA assumes significant importance in ensuring connectivity and throughput by simultaneously leveraging both low-band and mid-band channels. It has been recently reported [7, 35] that Ericsson, Nokia, Qualcomm, and mobile operators have successfully aggregated 6 CCs in a 5G SA network, achieving a peak downlink data rate exceeding 3.6 Gbps using only 5G low- and mid-band channels in FR1.

Despite the importance of CA, there have been limited studies of CA in real-world 5G deployments from the research community (see **§8**). CA adds significant complexity to the analysis of 5G network performance (see **§3**). For example, there are multiple modes ("peaks") in both 4G and 5G throughput distributions in Fig. 2 and Fig. 24. These can be attributed to the effects of CA, where "peaks" correspond to areas under coverage of multiple

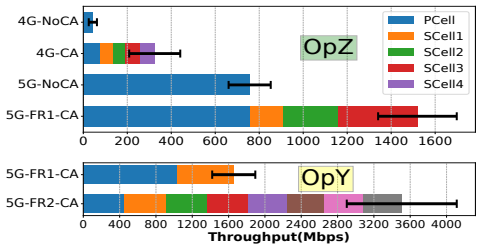
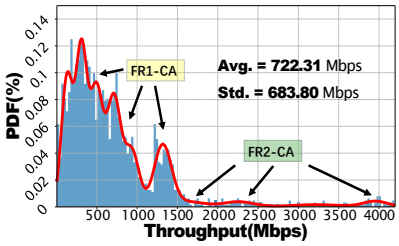
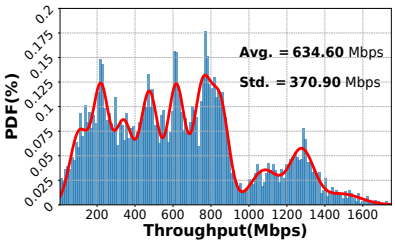


Figure 1: CA boosts both 4G and 5G networks throughput under ideal channel conditions. Each color indicates a component carrier (CC).



(a) OpY 5G throughput distribution.



(b) OpZ 5G throughput distribution.

Figure 2: Multi-modal distribution of 5G throughput during driving.

4G/5G bands/channels, thus multiple CCs with different combinations are available for CA. While previous measurement studies have more or less noted the highly variable [32, 33] and "multi-modal" nature of 5G throughput performance [48], our analysis reveals that carrier aggregation is one of the major contributors to such observed phenomena. If and when CA is activated, what and how many component carriers – and what band combinations – are used depends on various configurations and environmental factors, such as what bands/channels are configured and thus available in a given location, the channel conditions, and so forth. In addition, the capability of a 5G handset or user equipment (UE) also determines whether CA can be used for data transmission and what band combination is used. All of these affect the *user perceived 5G throughput performance* and application *quality of experience* (QoE). Using an XR (mixed/extended reality) application as an example (see **§3.3**), we show that while the significantly boosted 5G throughput enabled by 5G CA makes it possible to support applications with high bandwidth requirements, the application QoE may suffer due to high variability introduced by CA. Hence, in order to fully translate the throughput benefits brought by CA into the improvement of the application's QoE, it is imperative to take CA into account.

The goal of our paper is four-fold: First, we carry out a comprehensive measurement study of CA deployments over three major US operators in two large US cities. We map out the prevailing characteristics of the current CA deployments in 4G/5G networks, quantify the impact, and discuss challenges CA poses (**§3**). This is made possible via the use of a professional 5G measurement tool, Accuver XCAL [3], which allows a detailed collection of 5G New Radio (NR) PHY layer signals and RAN (radio access network) protocol messages; see **§2.2** for our measurement platform setup and methodology. Second, through careful experiments and in-depth data analysis, we dissect the complexity of 4G/5G CA configurations and identify the key factors that affect when and how CA is used (**§4**). The third and main goal of our paper is to develop a *CA-aware* deep-learning framework, dubbed *Prism5G*, for predicting 5G network throughput performance (**§5**) with the aim to aid application in QoE optimization. Through extensive evaluations, we demonstrate the benefits of *CA-awareness* in 5G throughput prediction (**§6**). Last but not least, we consider two use cases to illustrate how *Prism5G* can help applications enhance QoE performance (**§7**).

Contributions. We summarize our key contributions and major findings as follows:

• We conduct a timely and comprehensive measurement of CA deployment in commercial 5G networks (as well as 4G networks) in the US. To the best of our knowledge, this is the first *in-depth* study that considers the impact of CA on throughput analysis and prediction. We map out the CA deployments by all three major US carriers in two cities (and surrounding suburban areas and nearby highways), including the 4G/5G channels and combinations.

• Our study shows that diverse channels and channel combinations have been used to form 5G aggregate channels of exceeding 100 MHz, with up to 4 CCs in the mid-band and up to 8 CCs in the high-band (mmWave), resulting in peak throughput of more than 1.7 Gbps and 4.1 Gbps, respectively. While most of current 5G CA deployments are concentrated in urban areas, all three US operators are gradually expanding CA deployment and coverage.

• While 5G CA significantly boosts throughput, the complexity of CA poses new challenges in analyzing and predicting 5G performance. We demonstrate that not only is the activation and deactivation of CCs that induce drastic changes in 5G throughput in a short period of time, but the aggregated channel also exhibits far higher variability than when no CA is used. All of these have crucial QoE implications for (end-to-end) *bandwidth-adaptive* applications. Using (*scaled-up*) XR application ViVo [16] as an example, we show that CA worsens overall QoE metrics *comparatively*.

• The above findings call for the need of a *CA-aware* 5G throughput predictor that can more effectively aid applications in fast and adaptive decision making. Toward this end, we dissect the key factors that influence CA configurations and affect their performance. We demonstrate the need to capture features associated with individual CCs and the importance of accounting for the complex feature interplay in predicting 5G throughput performance.

• We propose a novel *CA-aware* deep learning framework, *Prism5G*, which models individual CCs, conditions them, and fuses them based on the CA state to accurately predict 5G throughput. It utilizes features that can be collected from UE. Evaluations using real-world 5G traces demonstrate the efficacy of Prism5G, with around 14% improvements over the state-of-the-art. It tracks the 5G throughput transitions well when CCs are added or removed.

• To demonstrate the utility of *Prism5G* in aiding adaptive applications in enhancing QoE performance, we consider two use cases: 1) We show that *Prism5G* can help ViVo to attain *near-optimal* QoE metrics. 2) Using MPC [50] as a representative adaptive bit rate (ABR) algorithm for video streaming, we show *Prism5G* enhances the average bit rates and reduces stall times considerably. It greatly improves the stall time *tail* performance, reducing the 95% percentile tail performance by 33 seconds (a ≈ 37% reduction). *Prism5G* consistently outperforms the other 5G throughput predictors.

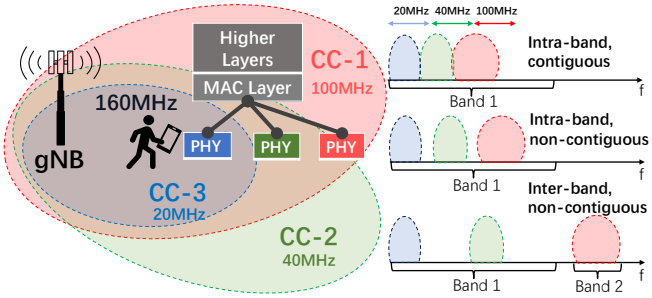


Figure 3: Illustration of carrier aggregation.

• We make measurement datasets, main codes, and other relevant artifacts publicly available[1].

2 BACKGROUND AND METHODOLOGY

This section introduces carrier aggregation (CA), followed by a description of our measurement setup and methodology.

2.1 A Quick Primer on CA

3GPP specifies a set of frequency bands within each of the low-band, mid-band, and high-band ranges that can be used to support 5G New Radio (NR). Each 5G band is designated with a number, prefixed with the letter "n" [45], e.g., n41 and n77 (C-band), both within the mid-band range. Similarly, 4G bands within the low- and mid-band ranges are also designated with a number [46], and we prefix them with the letter "b" to distinguish them from 5G bands. For each band, 3GPP also specifies what channel bandwidths can be supported, e.g., 15, 20, 40, 60, 70, 80, 90, or 100 MHz, and what subcarrier spacing (SCS) may be used, e.g., 15, 30, 60 or 120 kHz. We note that 4G bands use a fixed SCS of 15 kHz, and the maximum channel band is 20 MHz, compared to 100 MHz for 5G mid-band channels. For both 5G NR and 4G LTE, each band is specified to operate using either the TDD (Time Division Duplex) or FDD (Frequency Division Duplex) mode. In the TDD mode, both downlink (DL) and uplink (UL) data are transmitted using the same channel but in different time slots. Whereas in the FDD mode, a pair of channels (with the same frequency range) are dedicated to data transmissions, one for DL and one for UL.

At each 4G/5G base station, one or multiple channels (from the same or different bands) may be configured, depending on various factors, e.g., availability of spectrum, RAN capabilities, coverage, and bandwidth requirements of the service area. Each channel is often assigned a Physical Cell ID (PCI). The left panel in Fig. 3 depicts a base station with channels/bands/cells configured. We note that as radio bands have varying coverage ranges and different channel propagation characteristics, depending on the UE location, there may be one, two, or multiple channels/bands available. In a location where the UE is under the coverage of multiple cells, the cellular operator may opt to invoke CA by aggregating two or more channels to boost the throughput of data rates for the UE. Each channel is denoted as a *component carrier* and configured as a *serving cell* for CA. In the context of CA, the term (serving) cell is used interchangeably with CC.

[1]https://github.com/SIGCOMM24-5G-CA/artifact

Table 1: Overall Statistics of Collected CA Dataset.

Duration	Apr. 2023 - Jan. 2024			
Equipments	9 phones with 4 models			
Operators	AT&T, T-Mobile, Verizon			
# Freq. channels	4G: 86 5G: 44			
# CA combos	4G: 511 5G: 61			
Mobilities	Stationary, Walking, Driving			
Scenarios	Urban	Suburban	Beltway	Indoor
Cumulative data traces	140km 540min	180km 360min	470km 280min	10km 110min

As illustrated in the right panel of Fig. 3, CA may be performed using *contiguous* or *non-contiguous* channels with the same band – these are referred to as *intra-band contiguous* CA and *intra-band non-contiguous* CA, respectively. CA may also be performed using channels from *different* bands; this is called *inter-band CA*. In the latter case, the component carriers could be operating using TDD or FDD. In the case of 5G CA, each CC can also use different SCSs, e.g., 15 kHz or 30 kHz. For example, in the 5CC CA trial cited earlier [35], 3 TDD CCs and 1 FDD CC in the mid-band are used with the other 1 FDD CC in the low-band. In CA, one of the CCs is designated as the *primary* cell (PCell) over which radio resource control (RRC) messages are also transmitted, while all the other CCs are classified as *secondary* cells (SCells). SCells can be dynamically added or removed, depending on network conditions and other factors. CA is particularly important for 5G SA deployment as it is no longer anchored to 4G LTE, where low-band CCs can be aggregated to expand coverage. 3GPP specifies various band combinations that may be supported for NSA and SA CA operations. Lastly, we remark that CA is performed at the MAC layer where user data is multiplexed/demultiplexed across multiple CCs, as shown in Fig. 3. In a sense, NSA Dual Connectivity (DC) can also be viewed as a form of "CA" where traffic is split/merged between 4G LTE channels and 5G NR channels at the higher PDCP (Protocol Data Convergence Protocol) layer. We refer the reader to [2, 10] for more background and exposition on CA.

2.2 Measurement Methodology

To understand the CA deployment in real mobile networks, we conduct comprehensive measurements and summarize the statistical information of collected datasets in Table 1.

Operators, Locations and Mobility. Our measurement campaign mainly focuses on three major US mobile operators: AT&T, T-Mobile, and Verizon. We survey two large U.S. cities, covering their urban downtown, surrounding suburban areas, and major city beltways. Besides measurements conducted under driving mobility, which create a comprehensive coverage map, we also experiment with stationary mobility at various city hot spots, such as bus stops in line-of-sight to base stations, providing the baseline results under an ideal channel condition. In addition to outdoor measurements, we perform indoor measurements under walking mobility, a recognized challenging scenario for 5G. Overall, we acquire a rich dataset that provides us with a representative view of the current state of CA deployment. When showing the measurement results, we use OpX, OpY, and OpZ to obscure the operator names.

Table 2: 4G/5G channel allocation and combinations observed in our study. The frequency (abbreviated as Freq.) and bandwidth (abbreviated as BW) are measured in MHz. 4G bands are prefixed by the letter "b" and 5G by "n". The last column reports the number of observed channel combinations: the first number considers the ordering of SCells in the combinations, and the second one only counts the unique channel sets.

(a) Selected Channel and Band Allocation Observations												(b) Selected CA Observations		
Oper.	**Ch.**	**4G**							**5G**			**Channel Combo**	**Aggr. BW**	**Num.**
OpX	ID(Mode)	b12(FDD)	b14(FDD)	b29(FDD)	b2(FDD)	b66(FDD)	b30(FDD)	b46(TDD)	n5(FDD)	n77(TDD)	n260(TDD)	*4G* up to 5 CCs	(Up to) 90 MHz	270/162
	Freq.	700	700	700	1900	1700/2100	2300	5200	850	3700	39000	*5G* n77+n77	120 MHz	6/4
	BW	5,10	10	5	5,10,15,20	5,10,15,20	5,10	20	10	40,60,100	100	*5G* 8*n260	800 MHz	13/3
OpY	ID(Mode)	b13(FDD)	b5(FDD)	b4(FDD)	b2(FDD)	b66(FDD)	b48(TDD)	b46(TDD)	n5(FDD)	n77(TDD)	n261(TDD)	*4G* up to 5CCs	(Up to) 100 MHz	174/108
	Freq.	700	850	1700	1900	1700/2100	3600	5200	850	3700	28000	*5G* n77+n77	160 MHz	4/2
	BW	10	10	10,15,20	5,10,15,20	5,10,15,20	10,20	20	10	40,60,100	100	*5G* 8*n261	800 MHz	13/8
OpZ	ID(Mode)	b71(FDD)	b4(FDD)	b2(FDD)	b25(FDD)	b66(FDD)	b41(TDD)	b46(TDD)	n71(FDD)	n25(FDD)	n41(TDD)	*4G* up to 5 CCs	(Up to) 90 MHz	67/42
	Freq.	600	1700	1900	1900	1700/2100	2500	5200	600	1900	2500	*5G* n71+n41	120 MHz	7/7
	BW	5	10,15,20	5,10,15,20	5	5,10,15,20	20	20	15,20	20	20,40,60,100	*5G* n41+n71+n25+n41	180 MHz	6/2

Tool setup. We use the consumer smartphones as the 5G probes listed in Table 5 and ensure their firmware has been updated to the latest version as the 5G network configurations constantly evolve. These phones are tethered to a laptop running the professional network diagnostic tool XCAL [3]. This tool is used to access the chipset diagnostics and log data, including RRC control messages and precise radio frequency (RF) layer information. We adopt Iperf3 with multi-threads data transmission for our throughput measurement and set up an AWS EC2 instance (m5n.xlarge) as the remote server, which is able to provide a 4.1 Gbps baseline throughput and up to 25 Gbps burst throughput.

Methodology. There are numerous practical challenges, marked as ***[C]***, that cannot be neglected. These challenges may impact our results and, therefore, require specific remediation. ***[C1]*** We cannot directly control the carrier aggregation or easily select which channel becomes the component carrier. We resort to using the built-in function to force the technology and band that UE can use. For example, under the default coverage area of the band n71+n41 combination, we can force the UE to only use the channel within band n41 by entering the operator service code `*#2263#` for OpZ, while similar options are available on our professional tools XCAL [3] for the other operators. ***[C2]*** The continuous large-volume measurement traffic makes our UEs potentially compete for radio resources with other users and face throttling issues despite subscribing to unlimited data plans. To mitigate this, we mainly conduct experiments at midnight when fewer people are on the streets, and utilize multiple SIM cards to avoid potential data caps. We cross-validate the measurement results collected by running experiments multiple times on different days and filtering out outliers to ensure the results are representative. On the other hand, we have also collected the data at different times of the day to capture the time diversity, see the discussion in the Appendix B.2. ***[C3]*** The intensive data transmission with CA will quickly overheat the phone, leading to CA deactivation and a significant performance drop. We address this issue by engineering simple closed-loop liquid heat exchangers for cooling our smartphones and actively monitoring the phone's temperature during the measurement. ***[C4]*** The data recorded by multi-phones can lead to many out-of-synchronization problems and thus unfair performance comparison among different operators or CA configurations. To solve it, we place the phones side-by-side and master them on the same laptop; see Fig. 22 for example. Altogether, we carefully design our experiments, which provide a revealing snapshot of the state-of-the-art CA deployments.

3 MEASUREMENT & QOE IMPLICATIONS

This section presents the main findings from our measurement study, highlighting the key benefits and challenges posed by CA, with additional measurement observations provided in Appendix A. Using an XR application as an example, we also illustrate the impact of CA on application QoE.

3.1 CA Channel Allocation and Deployment

We start by discussing the CA channels, channel combinations, and CA deployment characteristics we have observed.

Diverse Channels and Channel Combinations. In our measurement data, we observed a total of 44 unique 5G channels and 86 unique 4G channels used for CA by the three major US operators. Most 5G channels come from the mid- or high-band (mmWave) ranges, operating in the TDD mode, with only a few operating in the FDD mode (in the below 2 GHz spectrum). In contrast, most 4G channels come from bands below the 2 GHz spectrum, operating in the FDD mode. The main difference lies in the channel bandwidth: The 4G channel bandwidth varies from 5 to 20 MHz. Whereas all 5G mid-band channels have a bandwidth of at least 20 MHz, most have a bandwidth of 40, 60 up to 100 MHz. In the case of mmWave channels, they are all 100 MHz. Table 2(a) provides representative samples of the 4G/5G bands/channels, and Table 6 in Appendix A.1 provides more detailed information. We see that many 5G channels share the same band with 4G channels – this is because cellular operators often "re-farm" their 4G spectrum for 5G services.

Via CA, the individual 4G/5G channels may be combined in various ways to form an *aggregated* channel of higher bandwidth. As selected combinations highlighted in Table 2(b), in the 4G networks of all three operators, up to 5 channels may be aggregated to yield an aggregated bandwidth up to 100 MHz – which is the maximum (allowed) channel bandwidth of a *single* 5G mid-band channel. In 5G networks, both OpX and OpY support up to 2 CCs of low- and mid-band channels (in FR1), with an aggregated bandwidth of up to 120 MHz and 160 MHz, respectively. Both these operators also support up to 8 CCs using mmWave channels (in FR2), yielding an aggregate bandwidth of up to 800 MHz. In contrast, OpZ supports

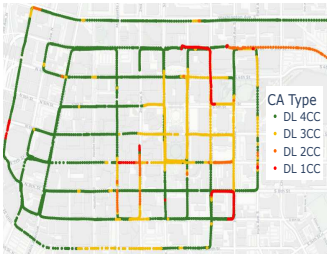


Figure 4: 5G CA deployment in an urban area.

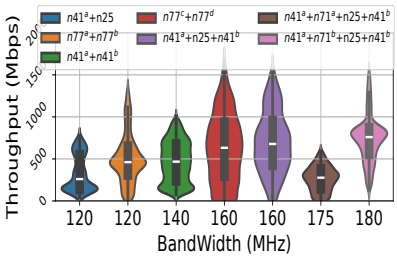
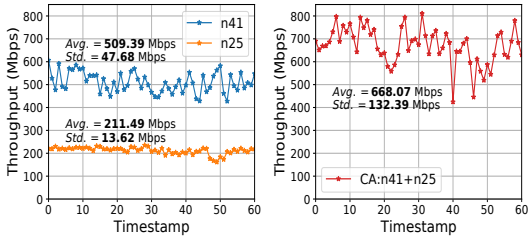


Figure 5: CA boosts Tput. The contour shows data distribution.



(a) Individual. **(b) CA: n25+n41.**

Figure 6: Throughput time series of n25 and n41 separately and aggregated in CA (n41+n25).

up to 4 CCs using channels from FR1 only, yielding an aggregated bandwidth of up to 180 MHz. See Table 7 in Appendix A.1 for sample channel combinations used in 5G CA.

4G/5G CA Deployment Prevalence. Our driving measurements in urban, suburban, and city beltway areas reveal that all three operators have widely deployed CA for both 4G and 5G networks, giving the mobile users a high likelihood of utilizing CA, as shown in Fig. 25. We observe that 4G CA covers almost the entire measurement area and provides seamless services, while 5G CA shows varying levels of prevalence across different areas, with averages of 24%, 44%, and 86% for OpX, OpY, OpZ, respectively. For the OpX and OpY, the deployment of mmWave 5G with up to 8 CCs aggregation is confined to densely populated areas. However, there is a notable ongoing expansion of their new 2 CCs aggregation in FR1. In contrast, OpZ has aggressively deployed 5G CA by re-farming their radio resources, thus providing more diverse CA options (more details can be found in Appendix A.1) and wider coverages not only in urban area but also suburban and city beltways. Therefore, we frequently use measurement data from OpZ in our following study.

Fig. 4 visualizes a sample spatial map of 5G CA deployment in an urban downtown area covering approximately $1km^2$, with measurements conducted along various streets. The color schemes indicate the number of CCs observed. We see that as the UE moves along a route, the number of CCs may fluctuate, with either a new CC added or an existing CC removed. Below, we will proceed to examine the implications of these observations.

3.2 CA Benefits and Challenges

As shown in Fig. 1 and Fig. 2, CA can significantly boost the throughput performance of both 4G and 5G networks. This is particularly the case in terms of *peak* 5G throughput under ideal channel condition: by aggregating 4 5G mid-band channel components, OpZ attains a peak throughput of 1.7 Gbps; with up to 8 mmWave channel components, OpY attains a peak throughput of 4.1 Gbps. In addition, we provide more measurement observations in Appendix A.2 regarding different mobility and scenarios. Fig. 26 shows that throughput can be significantly increased, up to multiple times, due to the use of CA while driving. Fig. 28 demonstrates that using FDD low-band as PCell in CA helps improve 5G signal connectivity while walking indoors, thereby providing higher throughput.

However, such significant performance gains are achieved at the expense of much added *configuration complexity* and *performance variability*, which make analyzing and predicting 5G network performance far more challenging. In 4G networks, because of the much narrow component carrier bandwidth, which varies from 5 to 20 MHz (in our data we observe that CCs of 20 MHz are most frequently used by all three operators), the observed aggregate 4G throughput is closely correlated with the number of CCs used in 4G CA. This is not true in 5G networks, due to more diverse channels/bands and much wider and varied channel bandwidth.

Fig. 5 shows the "violin" plots of the measured (aggregate) 5G throughput under 6 different 5G CA combinations from 2CCs to 4 CCs. We use the superscripts to distinguish different channels of the same band. Both with 2 CCs and an aggregated bandwidth of 120 MHz, the throughput performance of the n41^a+n25 combination differs vastly from that of the n77^a+n77^b combination (of two different n77 channels): the average throughput of the former is below 250 Mbps, about 1/2 of the latter, which is just below 500 Mbps. Both with an aggregate bandwidth of 160 MHz, the n77^c+n77^d combination and n41^a+n25+n41^b combination also exhibit quite different overall performance as indicated by the "fatness" of the contours, although they attain nearly the same peak throughput. Both with 4 CCs, the n41^a+n71^b+n25+n41^d combination exhibit significantly higher throughput than the n41^a+n71^a+n25+n41^d, whose SCell uses the different channel within the same band. With 4 CCs and slightly wider aggregate bandwidth, the n41^a+n71^b+n25+n41^d combination attains similar peak throughput as that of the n77^c+n77^d and n41^a+n25+n41^b combinations, but its overall throughput performance is more consistent, with higher average throughput than the 2 CCs and 3 CCs of 160 MHz. Compared to 4G CA, 5G CA in general introduces far higher performance variability, as noted in Fig. 2 and Fig. 24.

To better illustrate the performance variability introduced by 5G CA, in Fig. 6, we plot 60-second sample throughput trace segments of two 5G channels, n25 and n41, when both are used *alone* (i.e., no CA) as well as a sample throughput trace segment when they are combined as the 2 CCs aggregation (n41+n25). Data was collected at a fixed location with stationary UE, and the band was locked using the built-in function. First, we observe that the aggregate e throughput of n41+25 is not merely the sum of those of n41 and n25, sometimes at least 49.02% lower than the (theoretical) sum. In §4.3 we provide an explanation for this phenomenon. In fact, the channel characteristics and performance profile of an individual 5G channel can in general vary considerably from when it is used alone and when used in *different* CA combinations – this is because the configured power and MIMO layers may be altered (see §4.3 for an example using n41).

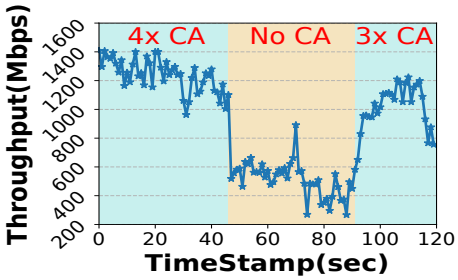


Figure 7: CA significantly changes throughput.

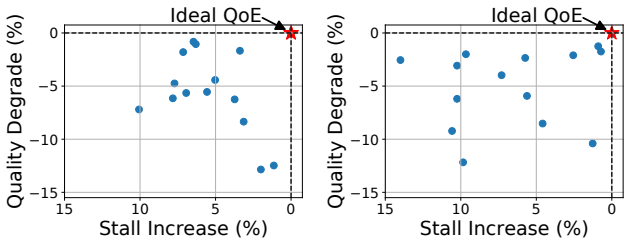
In Fig. 7, we plot a 120-second sample throughput trace segment with up to 4 CCs, where the CCs are dynamically added or removed as the user drives in a downtown area. We see that the addition and removal of CCs introduce drastic fluctuations in 5G throughput. For example, due to CC removal, around the time instance 46 sec, the throughput drops by about 1/2, from 1.2 Gbps to around 600 Mbps within a second or so; whereas from the time instance 90 to 94 sec, the throughput increases quickly from 550 Mbps to around 950 Mbps, due to the addition of 2 CCs.

Besides SCell's activation/deactivation, PCell may dynamically switch from one band to another, introducing additional complexities, such as transitions from TDD band to FDD band with altered power allocation. For an example of this occurring when a user moves from outdoor to indoor, see Appendix A.2. Furthermore, during the periods without addition or removals of CCs, the throughput of the aggregated channel often fluctuates more significantly than when no CA is used. All in all, while 5G CA provides cellular operators with the ability and flexibility to better utilize fragmented 5G channels by combining them to form aggregate channels of much higher bandwidth and boost 5G network throughput performance, it also poses new challenges. In the following, we explore the implications of 5G CA on application QoE.

3.3 Application QoE Implications

To illustrate the impact of 5G CA on application QoE, we use ViVo, an immersive XR application developed in [16] as an example. ViVo employs 3D point clouds to represent objects and the environment. To stream objects and their environment over networks, two key mechanisms are employed: a) ViVo first predicts the viewer's *field of view* 150 ms ahead to determine visible and unobstructed objects, shaping the 3D frame that must be delivered in the next 150 ms. b) ViVo adjusts the quality level (defined by point density) of the 3D frame to meet bandwidth constraints and the 150 ms delivery deadline. Similar to conventional video streaming ABR (adaptive bit rate) algorithms, ViVo uses past bandwidth measurements to estimate the available bandwidth in the next 150 ms [2]. Application QoE is therefore measured using two metrics: i) (average) quality level measured frame by frame; ii) (average) stall times, where a stall occurs if a 3D frame cannot be delivered within 150 ms.

We consider two scenarios. 1) ViVo over a 5G channel *without CA*: The average throughput is 355 Mbps with a standard deviation of 161 Mbps, and the peak throughput is 759 Mbps over various experiment runs. A segment of the throughput trace is shown in Fig. 6. 2) ViVo over an aggregate 5G channel with *4CCs* [3], where CCs may be dynamically added or removed: The average throughput is 700 Mbps with a standard deviation of 331 Mbps, and the peak throughput is 1732 Mbps. Refer to Fig. 7 for a segment of the throughput trace. In case 1), the maximum resolution of 3D point clouds and frame rate requires a maximal bit rate of 375 Mbps, slightly above the mean channel throughput. In case 2), to leverage the much higher aggregate channel bandwidth, we scale up the maximum resolution of the 3D point clouds and frame rate accordingly to 750 Mbps, also slightly above the mean throughput of the aggregate channel. In both cases, ViVo adapts to the fluctuating channel throughput by adjusting quality levels. To underscore the impact of 5G CA and motivate the need for *CA-aware throughput prediction*, we also consider an *ideal* version of ViVo, where the *actual* throughput in the next 150 ms interval (instead of estimated based on past measurements) is used.



(a) No CA (standard ViVo with bitrates up to 375 Mbps). **(b) 4CC CA (*scaled-up* ViVo with bitrates up to 750 Mbps).**

Figure 8: ViVo QoE under different CA conditions. Performance change is calculated via "ViVo-ViVo(ideal)."

Fig. 8 shows a number of representative results using various 5G traces with (a) no CA (i.e., case 1) and (b) with (upto) 4 CCs (i.e., case 2). Using the *ideal* ViVo as the baseline, the results are shown as the percentage of (average) quality degradation and the percentage of increases in (average) stall times. We see that without CA, there are multiple instances where both quality and stall times have degraded more than 5%. In the case of 4CC, the performance of most instances is visibly worse. While in several instances, the quality gradation is kept at 5%, this is achieved with significantly worsening performance in terms of stall times.

The above results underscore the fact that the current applications struggle to fully utilize the 5G network capabilities. It calls for advanced 5G performance prediction models to aid applications in more effective and adaptive decision-making, as in the case of *ideal* ViVo.

4 DISSECTING KEY FACTORS AFFECTING CA FOR THROUGHPUT PREDICTION

This section explores the complex interplay among various radio parameters that shape CA configurations and performance. Our goal is to develop an effective throughput prediction algorithm, so our focus lies on UE-collectible parameters or "features" (cf. Table 3), e.g., via Android APIs [6]. We use OpZ as the primary mobile

[2]In contrast to video-on-demand ABR algorithms, which plan quality levels for video chunks seconds ahead because of using a large buffer, ViVo's quality adaptation algorithm operates at a much shorter time scale (in hundred ms level), making frame-by-frame decisions with a "shallow" buffer.

[3]We note that the 4 CCs can aggregate up to 180 MHz bandwidth and do not equate to a fourfold increase in throughput as each channel has a different bandwidth (see Fig. 5 and Table 2).

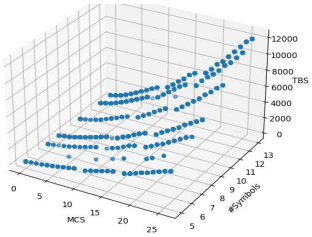


Figure 9: The mapping among transport block size (TBS), MCS, and resource element number, when using 2 MIMO layers.

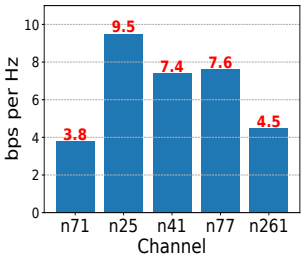


Figure 10: Selected 5G channel efficiency with good channel condition (CQI>12).

operator, given its extensive CA coverage and diverse channel combinations.

4.1 Key 5G PHY Radio Parameters

From the physical layer (PHY) perspective, 5G throughput (theoretically) depends on several key parameters. With CA, aggregated throughput is the sum of individual CC throughputs, making the number of CCs crucial. For each CC, its throughput depends on several key factors: the maximum channel bandwidth which determines the (maximum) configured resource blocks (RBs) and thus affects the number of RBs that may be allocated to each UE (in the frequency domain); the number of symbols allocated to the UE *per slot* (in the time domain); and the number of MIMO layers (#Layers) used. The frequency and time domain radio resource allocation together yield the number of resource elements (REs) allocated to the UE per slot. The number of bits carried in each RE and thus the total number of bits carried in each slot – referred to as the *transport block size* (TBS) – is determined by the modulation and coding scheme (MCS). Hence, the throughput (Tput) of each CC is a function of these parameters, namely, $Tput = f(MCS, \#RE, \#Layers)$ (see Appendix B for the theoretical calculation of PHY throughput). The sample results in Fig. 9 show the relationship among the PHY throughput, measured in terms of TBS (bits), the number of symbols allocated per slot, and employed MCS, where the number of MIMO layers is fixed to 2.

Both MCS and the number of MIMO layers hinge on channel conditions (as well as other factors such as the amount of buffered user data). UE periodically feeds back channel state information, such as channel quality indicator (CQI) and rank indicator (RI), to aid the base station in deciding on the MCS and the number of MIMO layers for data transmissions. CQI itself is a function of the reference signal received power (RSRP) and quality (RSRQ), signal-to-noise-ration (SINR), etc., measured at the UE.

As channels in different bands have distinct radio propagation characteristics and may be subject to varying environmental factors, not all channels/bands are equal. As an example, Fig. 10 shows the spectrum efficiency (defined as bits per second (bps) per Hz) for five channels from the low-, mid- and high-band ranges, measured under an ideal scenario (the best channel condition using the highest MCS and full RB allocation). Whether CA is invoked (configured) – and when it is invoked, the number of CCs used – also hinges upon the channel conditions (among other factors).

4.2 Need for Modeling Each Channel/Band

We argue the need for capturing the channel characteristics or "features" of individual component carriers separately, especially when they come from different bands, in order to predict their throughput accurately. We use the Synchronization Signal RSRP (SS-RSRP) as an example feature to illustrate the point. We consider two cases involving CA with two intra-band CCs vs. two inter-band CCs: 1) one n41 CC of 100 MHz bandwidth (PCell) and another n41 CC of 40 MHz (SCell); and 2) one n41 CC of 100 MHz bandwidth (PCell) and another n25 CC of 20 MHz.

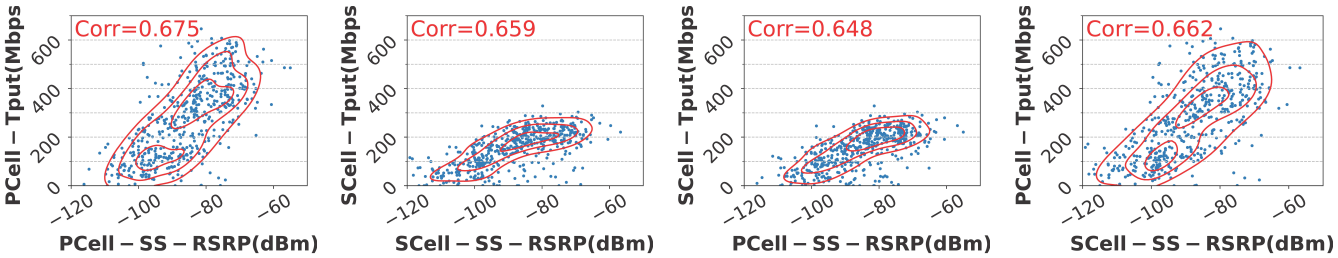
In Fig. 11 and 12, we present the correlation (measured by the Pearson coefficient) between reported RSRP and observed throughput of PCell and SCell for intra- and inter-band scenarios. Fig.11 (a) and (b) show the correlation between PCell RSRP and throughput in the case of intra-band CA using 2 n41 CCs. In Fig. 12 (a) and (b), we show the same results in the case of inter-band CA using 1 n41 CC and 1 n25 CC. In both cases, there are strong correlations (Corr > 0.6) between the throughput of each CC and its RSRP. In contrast, in Fig. 11 (c) & (d) and respectively Fig. 12 (c) & (d), we show the correlation of the RSRP of one CC with the throughput of another CC. We see that in the case of the intra-band CA, the correlation is still above 0.6, whereas in the case of the inter-band CA, the correlation has dropped significantly, to only 0.5 in terms of PCell-RSRP and SCell-Tput and to 0.55 in terms of SCell-RSRP and PCell-Tput.

In Fig. 13a and Fig. 13b, we further compare intra-band CA vs. inter-band CA by examining the correlation between the RSRP of the PCell and that of SCell. We see that for intra-band CA, the RSRPs of two CCs are highly correlated and track each other very well over time. Whereas for inter-band CA, the RSRPs of two CCs are not strongly correlated over time. Therefore, for inter-band CA, simply using the reported RSRP of one CC (say, the PCell) to predict the throughput of another CC (e.g., one of the SCells) may lead to a suboptimal result.

4.3 Need for Accounting for Complex & Dynamic Feature Interplay

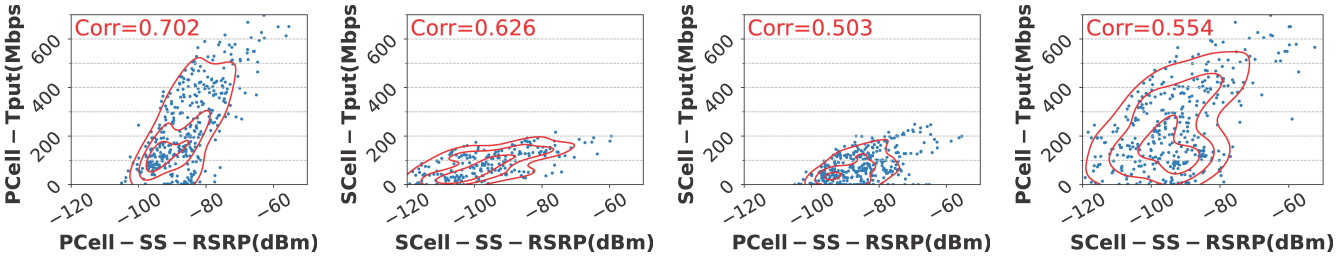
In the above, we see that the channel's RSRP is strongly correlated with its observed throughput. However, the RSRP of a channel alone is not sufficient to predict its throughput. We will use examples to illustrate. The examples demonstrate the need to account for the complex and dynamic interplay among various radio channel features, and thus make the case for sophisticated deep learning methods for *CA-aware* throughput prediction.

We first consider and compare the throughput of a channel when CA is not invoked vs. that of the same channel when it is used as part of 3CC CA. Fig. 14 shows the measured throughput results for an n25 channel *with and without CA at the same location*. In both cases, the reported RSRP and CQI of the channel are similar (and the number of allocated RBs is also similar). We see that there is a significant difference in observed throughput: without CA, the throughput of the n25 channel is above 200 Mbps on average, whereas with CA, its throughput drops to only slightly above 100 Mbps on average. This is due to the fact that with CA, the number of MIMO layers used for the n25 is reduced from 3 to 1 (likely due to the reduced transmission power for the n25 channel by the base



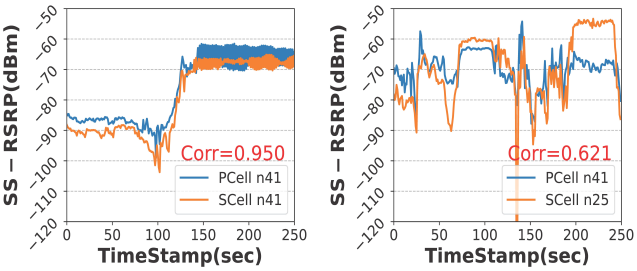
(a) PCell RSRP v.s. PCell Tput. (b) SCell RSRP v.s. SCell Tput. (c) PCell RSRP v.s. SCell Tput. (d) SCell RSRP v.s. PCell Tput.

Figure 11: The intra-band (PCell n41, SCell n41) correlation between CC's received signal strength and throughput.



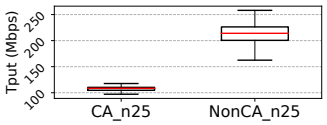
(a) PCell RSRP v.s. PCell Tput. (b) SCell RSRP v.s. SCell Tput. (c) PCell RSRP v.s. SCell Tput. (d) SCell RSRP v.s. PCell Tput.

Figure 12: The inter-band (PCell n41, SCell n25) correlation between CC's received signal strength and throughput.



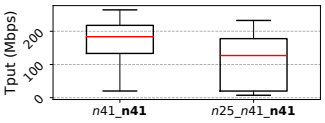
(a) PCell (n41) v.s. SCell (n41). (b) PCell (n41) v.s. SCell (n25).

Figure 13: Correlations of received signal power between PCell and SCell.



	CA n41_***n25***_n41	NonCA ***n25***
RSRP	-68.1dBm	-69.6dBm
CQI	12.2	12.1
MIMO	**1**	**3**
#RB	102.4	103.4
Total Tput	862 Mbps	212 Mbps

Figure 14: Throughput of the same channel *with* and *without* CA, using *n25* as an example.



	CA n41_***n41***	CA n25_n41_***n41***
RSRP	-97.1dBm	-99.1dBm
CQI	8.7	9.2
MIMO	2	2
#RB	**69.3**	**55.1**
Total Tput	433 Mbps	565 Mbps

Figure 15: Tput of the same channel under different CAs, using the second *n41* channel as an example.

station). In Fig.15, we compare the throughput of the same channel (an n41 channel of 40 MHz) that is used in different CA band combinations: 1) n41-n41 intra-band CA (with the first n41 channel of 100 MHz being the PCell) and 2) n25-n41-n41 inter-band CA (with the n25 channel being the PCell). We see that the throughput of this n41 channel (of 40 MHz) differs significantly in these two cases, despite the fact that it is used as the SCell in both cases, and the RSRP and CQI are also similar. Furthermore, the number of MIMO layers used is also the same. The difference in the observed throughput can be attributed to the number of RBs allocated. With the other CCs having 120MHz bandwidth, the additional SCell may be throttling out in the service busy area. The above examples show the importance of taking into account the diverse band combinations as well as the complex and dynamic interplay among various channel parameters. Therefore, *CA-awareness* is critical to 5G throughput prediction.

5 CA-AWARE TPUT PREDICTION MODEL

This section introduces *Prism5G*, a novel *CA-aware* deep learning framework for 4G/5G throughput prediction. It models and conditions individual CCs, fusing them based on CA state for accurate 5G throughput prediction (see Fig. 16), using the UE-collectible features (see Table 3).

5.1 Overview of *Prism5G*

Challenges. The widespread adoption of CA in 4G/5G has significantly improved achievable throughput for end users. However, accurately modeling the performance of throughput on the UE side becomes particularly challenging in the presence of CA, especially within the realm of 5G. **(1) Heterogeneity:** The diversity of 5G channels, and their combinations, exhibit distinct characteristics. Additionally, the availability and configuration of these channels may vary across locations and network deployments. **(2) Complexity:** The intricate interplay and correlations among channels within a CA configuration escalate the complexity, rendering the

Table 3: Features used for the machine learning with illustrations in Table 12.

Hardware	5G Modem Model
Signaling	Radio Resource Control CA Events
PHY Info [per CC]	Band/Channel Info; ssRSRP;ssRSRQ; SINR; CQI; BLER; (Optional): #RB; #Layers; MCS;
App	Historical Tput

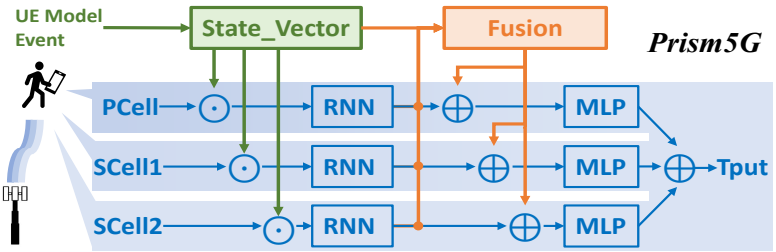


Figure 16: CA-aware ML model for Tput prediction.

explanation of the ML model challenging. **(3) Data Scarcity:** The laboriousness of measurements and commercial constraints impede telecom companies from scaling the open-sourced datasets obtained from their commercial networks, reducing the data available for training and evaluating ML models. **(4) Lightweight:** The limited compute resources on mobile UE and need for real-time inference necessitate lightweight ML models.

Our Design. To address these challenges, we design a deep learning framework, denoted as *Prism5G*, which exhibits *adaptability*, *flexibility*, and *explainability* for predicting 5G network throughput. Distinguishing itself from existing approaches [4, 28, 32] that blindly predict overall throughput, *Prism5G* endeavors to harness the above measurement observations and domain knowledge of 5G networks to explicitly account for the impact of the CA mechanism.

Fig. 16 shows the overall design of *Prism5G*. It consists of three core principles: (1) modeling of each CC (blue), (2) monitoring the signaling events (green), and (3) fusion learning for the interactive correlation (orange). Specifically, *Prism5G* adopts a weights-shared neural network to predict the future throughput of each CC. These individual predictions are then aggregated to obtain the overall throughput. Such explicit consideration of CA enables *Prism5G* to make predictions at a fine-grained level (per CC), thereby achieving a certain degree of flexibility and explainability compared to directly modeling the overall throughput. Meanwhile, *Prism5G* takes the signaling events and distinct capabilities of different chipset modems as domain knowledge for each prediction. This knowledge is transformed into masks that explicitly adjust the states of carrier components. By incorporating the domain knowledge, *Prism5G* achieves faster adaptation to network environment changes and reduces complexity by eliminating the need for indirect learning from the physical layer. Finally, *Prism5G* also explicitly considers the interaction between CC under different channel combinations. Collectively, these designs empower *Prism5G* with the capability to effectively and accurately predict the throughput performance of 4G/5G networks. Detailed information for each module is provided below.

5.2 Model Module Explained

Per CC Modeling. Let C denote the total number of carrier components and $\mathbf{X}_c$ denote the features of each carrier component c. These features (see Table 12) include information from the past T time steps, such as throughput and physical channel quality, i.e., $\mathbf{X}_c = [X_c^{t-T+1}, ..., X_c^{t-1}, X_c^t] \in R^{(C,T)}$. Based on these input features, we employ an RNN module to predict the future throughput of each carrier component: $h_c = RNN_{\theta_1}(X_c)$, where θ_1 denotes the trainable parameters of the RNN layers. The RNN modules share weights to leverage shared knowledge and reduce the number of parameters, thereby decreasing the overall complexity of *Prism5G*. The type of RNN module is configurable, and we use LSTM in our experiments. Although we use it as our building block when developing *Prism5G*, the design of *Prism5G* does not require a specific DNN architecture. In other words, it can be easily replaced by other similar (or more advanced) building blocks, such as transformers.

CA Event Monitoring. To capture the dynamics of the channel combination over time, *Prism5G* translates the signaling control messages obtained from the Radio Resource Control (RRC) protocol into a binary mask vector, denoted as $\mathbb{I} \in R^{(C,T)}$. This vector is responsible for activating and deactivating RNN modules (i.e., carrier components): $\mathbf{X}_c^{'} = \mathbf{X}_c \odot \mathbb{I}$. Furthermore, in order to provide the fusion learning module with a richer context of the current channel combination, *Prism5G* utilizes an embedding layer to transform the sparse binary mask vector $\mathbb{I}$ into a dense embedding $\mathbf{E}$.

Fusion Learning. This module fuses the RNN hidden states of all carrier components to extract the interplay and correlations among different channels, taking into account the current channel combination condition: $h_f = Fusion_{\theta_2}([h_1, \ldots, h_c, \ldots, h_C, \mathbf{E}])$, where θ_2 represents the trainable parameters of the fusion learning module.

Aggregated Throughput Prediction. *Prism5G* first aggregates the RNN hidden state h_c of each component carrier and the overall channel correlation h_f, i.e., $h_c^{'} = h_c + h_f$. This aggregated information is then inputted into an MLP module to predict the future throughput of each carrier. The overall throughput for an end user is obtained by aggregating all the predicted throughput for each carrier. The whole process is presented as $y_{pred} = \sum_{c=1}^{C} MLP_{\theta_3}(h_c^{'})$, where y_{pred} denotes the predicted aggregated throughput. We jointly train all the aforementioned modules. The optimal parameters $\Theta = [\theta_1, \theta_2, \theta_3]$ of *Prism5G* are obtained by minimizing the prediction errors: $\Theta^{\star} = \min_{\Theta} \mathcal{L}(y_{pred}, y_{true})$.

6 EVALUATION

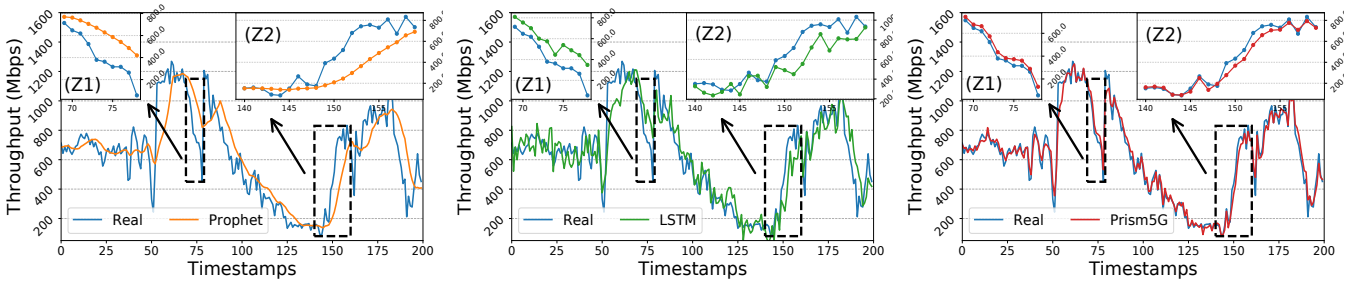
This section evaluates *Prism5G*'s performance, highlighting a 14.0% reduction in root-mean-square-error (RMSE) compared to the best baseline, with efficient training and inference times. Notably, *Prism5G* displays remarkable adaptability to network changes and strong transferability across diverse settings.

6.1 Experiment Setups and Implementations

Datasets. We extract and process self-collected datasets at 10ms and 1s time granularity, obtaining a total of 6 sub-datasets for machine learning evaluation, each comprising thousands of valid data records. These sub-datasets encompass diverse operators, channel combinations, and scenarios. Each scenario contains 10 traces with

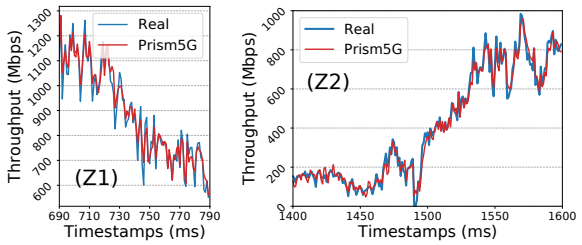
Table 4: Performance results of *Prism5G* and selected baselines measured in RMSE, lower values being better.

Dataset	**Short** (10ms time scale with 100ms prediction horizon)						**Long** (1s time scale with 10s prediction horizon)					
	Prophet	LSTM	TCN	Lumos5G	Ours	Improv. (%)	Prophet	LSTM	TCN	Lumos5G	Ours	Improv. (%)
OpX (Walking)	0.391	0.223	0.237	0.282	**0.188**	15.70%	0.317	0.199	0.200	0.206	**0.187**	5.47%
OpX (Driving)	0.416	0.339	0.358	0.339	**0.283**	16.52%	0.333	0.223	0.230	0.228	**0.200**	10.31%
OpY (Walking)	0.362	0.226	0.229	0.340	**0.195**	13.72%	0.327	0.211	0.233	0.219	**0.192**	9.00%
OpY (Driving)	0.434	0.327	0.360	0.349	**0.274**	16.21%	0.381	0.289	0.290	0.295	**0.260**	10.03%
OpZ (Walking)	0.408	0.303	0.325	0.301	**0.240**	20.27%	0.376	0.276	0.291	0.265	**0.228**	13.96%
OpZ (Driving)	0.607	0.462	0.451	0.448	**0.352**	22.00%	0.451	0.342	0.337	0.325	**0.277**	14.77%



(a) Prophet. **(b) LSTM.** **(c) *Prism5G*.**

Figure 17: Visualized prediction results at *1-sec* time scale. Two transition point areas are marked as Z1 and Z2.



(a) Throughput drop. **(b) Throughput boost.**

Figure 18: Prediction at the transition point area (Z1 and Z2) with a *10-ms* time scale.

300 to 600 data samples per trace. In Appendix C, Table 11 summarizes the statistical information of these sub-datasets, and Table 12 illustrates the description of each data field. Each trace uses a moving window to create data pairs, consisting of a historical window and a future window. Those data pairs are then split into training, validation, and test sets based on a specified ratio.

Baseline Setups. We compare *Prism5G* with six baseline models, which can be categorized into three groups: (1) Statistic-based time-series forecasting: Prophet [44]; (2) Classical machine learning (widely used for throughput prediction due to their explainability): GDBT [32], RF [4]; (3) Deep learning-based: LSTM [28], TCN [9], and Lumos5G [32] [4]. For all the evaluations, we set the input and output sequence length to 10, signifying the 100ms (or 10s) prediction horizon, depending on the dataset time granularity. We use root-mean-square-error (RMSE) as the loss function and report the optimal model based on its validation set performance. For the concrete training and evaluation strategy for Prophet and classical ML, refer to Appendix C.1.

Runtime. Compared with LSTM, *Prism5G* introduces an additional 34.1% training time on average. The training duration typically ranges from 5 to 30 minutes, depending on the dataset and hyper-parameter selection. As for inference, although *Prism5G* incurs an extra 23.2% time on average, yet remains below 1 ms per sample, well below decision time (10ms for ViVo and 1s for ABR in §7).

6.2 Prediction Results and Comparisons

Overall Accuracy. Table 4 shows the overall performance of *Prism5G* and baselines performance measured in RMSE on the collected 5G traces. The experiments are conducted in both the short period using 10 ms time scale with 100 ms prediction horizon, and the long period using 1 s time scale with 10 s prediction horizon. We can see that: (1) *Prism5G* is consistently superior to the current best baseline, with an average of 14% and a maximum of 22% reduction in RMSE. (2) *Prism5G* outperforms other algorithms in both time scales with an average of 17.4% and 10.6% reduction in RMSE for both the short and long time scales, respectively. (3) Purely time series prediction algorithms (e.g., Prophet) do not work well on 5G datasets, having the highest RMSE across all the other algorithms for nearly all datasets.

We visualize the prediction results in Fig. 17 by plotting the first predicted point in the horizon window. For simplicity, we only select two representative baselines (Prophet and LSTM) for comparison with *Prism5G*. We delve into two critical instances: area Z1 is marked by a significant throughput drop due to SCell deactivation and worse channel quality, and area Z2 is characterized by a notable performance boost due to SCell activation and better channel quality. We can notice that in Z1, Prophet and LSTM are overestimating the throughput, while underestimating it in Z2. *Prism5G* has the closest prediction to the real throughput in both

[4]We mainly consider the Lumos5G's model architecture (i.e., Seq2Seq) for comparison, as some user-context features designed for mmWave 5G, such as panel angle, user movement direction, and orientation, are not suitable in most of the non-line-of-sight scenarios happened in low- and mid-band.

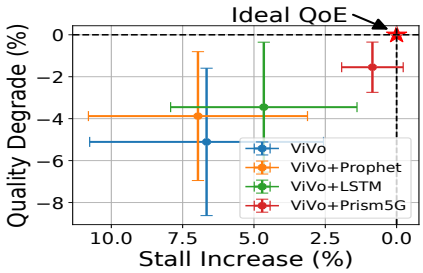


Figure 19: *Prism5G* improves ViVo.

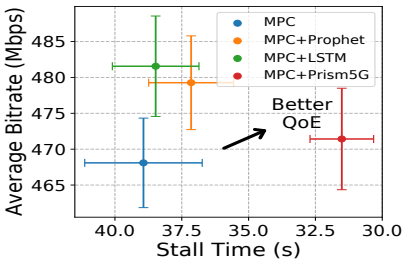


Figure 20: ABR video streaming.

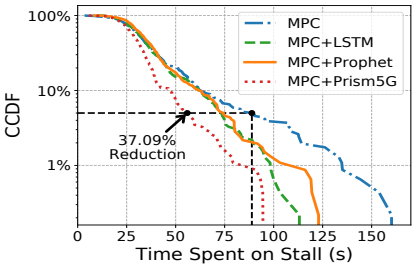


Figure 21: Tail performance.

cases. We refer the reader to Appendix C.2 for additional results, where Fig. 33 and Fig. 34 show each CC's predicted results by *Prism5G*, where each cell is well modeled.

Transition Point. We also visualize the 10ms short-period prediction results of the representative transition point areas Z1 and Z2 in Fig. 18. This allows us to see that *Prism5G* exhibits a faster reaction at the transition points and closely matches the real throughput. In turn this is indicative of its availability to assist application decisions at a finer time granularity.

Ablation Study. We perform an ablation study to demonstrate the necessity of two key mechanisms in our model: (1) the state trigger mechanism and (2) the fusion mechanism. As shown in Table 13 in Appendix C.2, without the state trigger mechanism, the average and maximum RMSE increase by 5.3% and 7.1%, respectively. Similarly, without the fusion mechanism, these increase by 6.2% and 9.5%.

Generalizability. Expanding on our previous evaluation, where each trace contributes to both training and test sets, we delve deeper into assessing the generalizability of the proposed *Prism5G*. We use the sub-dataset of OpZ with the walking mobility at 1s time scale for this study. We first evaluate the *Prism5G* on the same route but different runs. To do this, we split the dataset based on the traces instead of splitting them as in the previous evaluation. Table 14 showcases the consistent superiority of the *Prism5G*, achieving an average 9.4% lower RMSE compared to the best baselines. Additionally, we extend our evaluation to include new traces collected on different routes, not part of the original sub-dataset. We observe that the *Prism5G* outperforms all baselines, maintaining strong performance with an average RMSE 12.5% lower than the baselines.

In summary, *Prism5G* outperforms the current state-of-the-art algorithms and has the highest accuracy for predicting 5G throughput for the aggregate and individual cells. It quickly adapts to transition points when an SCell is activated or deactivated with high accuracy.

7 *PRISM5G* USE CASES

This section employs two use cases to showcase *Prism5G*'s efficacy in enhancing application QoE through smart decision-making. We also compare *Prism5G* with Prophet and LSTM.

XR Immersive Content Delivery. Recall from **§3.3** that ViVo employs a simple mechanism to estimate future network bandwidth based on past data and use it to decide on the quality level of 3D frames that are 150ms ahead. The decisions are made frame-by-frame at a rather fast time scale (10's ms). We replace this simple mechanism with *Prism5G*, which predicts 5G throughput at a fast time scale (10s ms) with a short horizon (100s ms). We denote this as ViVo+*Prism5G*. We also compare with ViVo+Prophet and ViVo+LSTM. Experiments are done using 2300+ 5G traces with (up to) 4CCs with average aggregate throughput from 223.86 to 627.69 Mpbs. We use the scaled-up ViVo application as in **§3.3**. The results are shown in Fig. 19. We see that ViVo+*Prism5G* attains *near-optimal* performance, with QoE results very close to those of the *ideal* ViVo. In contrast, ViVo+LSTM also yields improved QoE performance, but it is far from near-optimal. While ViVo+Prophet also improves the average quality levels, but the improvements come at the expense of slightly worsening average stall time performance.

UHD Video-on-Demand Streaming. In contrast to ViVo, ABR algorithms used in video streaming make decisions at a longer time scale (at the level of chunk length, typically a few seconds long). In other words, decisions are made to prefetch several video chunks ahead at certain quality level, and a large client buffer is used to store prefetched data to accommodate network bandwidth fluctuations. We use MPC [50], a widely used throughput-based ABR algorithm, as an example[5]. MPC uses a simple predictor based on Harmonic Mean which estimates the future bandwidth using data from the recent past. We replace this predictor with *Prism5G* to predict future 5G throughput over a time horizon of 10 seconds at the time scale of seconds. We emulate streaming of 16K videos over 5G networks. Each video is encoded using H.264 codec in multiple quality levels: [1.5, 2.5, 40.71, 152.66, 280, 585] Mbps, corresponding to resolutions [360p, 480p, 2K, 4K, 8K,16K]. Evaluations are performed using the 5G CA traces collected. The results are shown in Fig. 20.

We see MPC+*Prism5G* yields significant QoE improvements over MPC: while it increases the average bit rate modestly from below 468 Mbps to around 472 Mbps (*i.e.,* a 0.71% improvement), it reduces the average stall time significantly from around 39 seconds to below 31 seconds (*i.e.,* a 19.06% reduction). In contrast, while MPC+Prophet and MPC+LSTM increase the average bit rate considerably (*i.e.,* by 2.5%), these improvements are accompanied by only slight (*i.e.,* 2.8%) reductions in the average stall time. This is largely because during the transition periods when CCs are deactivated, both Prophet and LSTM significantly over-estimate 5G throughput during these periods (cf. Fig. 18). While such over-estimates increase the overall bit rates, they can induce significant stall events. Furthermore, what the *average QoE statistics* do not show are the significant improvements afforded by *Prism5G* in stall time *tail performance* (especially during the transition periods when

[5] *Prism5G* can be directly incorporated in other throughput-based ABR algorithms. For buffer-based ABR algorithms such as BOLA [43], hybrid or other ABR algorithms, *Prism5G* can be used to simulate and predict future buffer occupancy or other system state that is critical in decision making.

5G throughput drops significantly, due to CC de-activation). As shown in Fig. 21, MPC+*Prism5G* improves 99%, 95%, and 90% stall time tail performance by 50.8, 33.0, and 16.0 seconds, respectively, far better than those of MPC+Prophet and MPC+LSTM.

8 RELATED WORKS

Deployment of commercial 5G services around the world since 2019 has attracted a flurry of measurement studies of 5G networks. These studies have revealed crucial insights concerning coverage, latency, throughput, and application performance [8, 11, 12, 14, 17, 26, 31–34, 36, 37, 39, 40, 47, 49]. In the following we will focus our attention on 4G/5G carrier aggregation and 3G/4G/5G throughput prediction related research.

4G/5G Carrier Aggregation & Measurement. Although 3GPP has released the technical specification of CA in 5G networks since 2019 [1], the commercial deployment of 5G CA is still in its early stages, with fairly limited literature studies. Lin *et al.* [24] discussed the importance of CA from the technical specification perspective, particularly in the context of operators' co-construction and sharing. Liu *et al.* [27] conducted a measurement study on the general 5G experience with three US operators and proposed a patch solution for configuring radio resource control and making better choices to select service cells. Li *et al.* [23] examined the signaling process of CA access from the control plane perspective primarily using commercial 4G measurement datasets, as commercial 5G-CA on mid-band had not been widely deployed at that time. Narayanan *et al.* [33] performed measurements of mmWave 5G (FR2) deployments but only studied the impact of CC number on peak throughput. Fezeu *et al.* [11] further study video streaming performance in the context of different frequency bands and examine the impact of channel variability through scaled variability metrics. Other works [15, 18, 38, 40, 48] primarily focused on investigating the capacity of 5G channels and achievable performance with limited in-depth discussion toward CA. Additionally, telecommunication companies have conducted preliminary CA measurements [7, 35] but only in controlled trial environments rather than on large-scale commercial deployments. Unlike existing studies, we conduct a timely and comprehensive measurement of CA deployment in commercial 5G networks and 4G networks in the US, with the consideration of CA impacts on throughput analysis and prediction.

3G/4G/5G Throughput Prediction. There is a large amount of literature on throughput prediction [21, 22, 29, 41, 42, 51] for 3G/4G networks, utilizing both machine learning-based and analytical approaches. 5G throughput prediction is notably more challenging than 3G/4G due to diverse bands, complex technologies, and various other factors involved. Lumos5G [32] employed GDBT and Seq2Seq to predict mmWave 5G throughput based on UE-side contexts. Mei *et al.* [28] employed the LSTM model to capture the temporal patterns of bandwidth evolution. Minovski *et al.* [30] trained ML models in LTE networks to predict throughput using lower-layer radio-related metrics. They further fine-tuned these models for non-standalone 5G networks. A hybrid prediction method is proposed in [25], which utilizes an ARMA time series model for intra-cell bandwidth prediction and a Random Forest (RF) regression model for cross-cell bandwidth prediction. However, previous studies have not explicitly integrated CA into ML throughput prediction models, making them inadequate for CA-enabled 5G networks. In contrast, *Prism5G* stands out as a flexible framework that accommodates various DNN modules. It incorporates measurement observations and domain knowledge of CA into its design, enabling accurate 5G throughput prediction with per-component carrier modeling and accounting for carrier interactions.

9 CONCLUSIONS

We have carried out a timely in-depth study of carrier aggregation (CA) in commercial 5G (and 4G) networks. Through comprehensive measurement-based analysis, we not only demonstrate how CA can significantly boost 5G network throughput performance, but also illustrate the new challenges posed by CA and their implications on application QoE performance. Our findings lead us to pursue the development of a *CA-aware* 5G throughput prediction framework that can effectively aid applications in fast and adaptive decision-making. To this end, we have identified the key factors influencing whether or when CA is deployed and what band combinations are used. Building on these insights, we have developed *Prism5G*, a novel deep learning predictor for 5G throughput prediction which explicitly considers the complexities introduced by CA, thus is *CA-aware*. Using real-world 5G traces, we have extensively evaluated the efficacy of Prism5G and compared its performance with existing 5G throughput prediction algorithms. Our results demonstrate that the *Prism5G* outperforms the state-of-the-art algorithms by over 14% on average in terms of prediction error. Through two use cases, we further demonstrate the utility of *Prism5G* in aiding applications (and networks) to optimize QoE performance.

Future Directions. The performance of CA under multi-user competition scenarios has not been well studied and requires further exploration. We plan to evaluate the trade-offs of adapting various learning models, such as transformers, to the *Prism5G* framework. Future efforts will also focus on implementing a real-world benchmark platform with downstream applications for mobile throughput prediction evaluation and optimizing system overhead.

Ethical Considerations. The measurement study was conducted by the research team in compliance with wireless carriers' agreements. No human subjects were involved, nor was any personally identifiable information collected.

ACKNOWLEDGMENT

We thank our shepherd Prof. Chunyi Peng and other anonymous reviewers for their feedback and suggestions. We thank Muhammad Iqbal Rochman and Prof. Monisha Ghosh for their assistance in data collection and for valuable insights and feedback on this research. We also thank Faaiq Bilal and Jiaxiang Tang for their help during the revision stage. This research was supported in part by NSF under Grants 1915122, 2128489, 2154078, 2212318, 2220286, 2220292, and 2321531, as well as a Cisco Research grant and an InterDigital gift.

REFERENCES

[1] 3GPP. 2019. 3GPP TS 38.101-2 version 15.4.0 Release 15 - Part 2: Range 2 Standalone. https://www.etsi.org/deliver/etsi_ts/138100_138199/13810102/15.04.00_60/ts_13810102v150400p.pdf
[2] 3GPP. 2022. Carrier Aggregation explained. https://www.3gpp.org/technologies/101-carrier-aggregation-explained. Accessed: 2023-06-09.
[3] Accuver. 2021. XCAL. https://www.accuver.com/sub/products/view.php?idx=6&ckattempt=2. Accessed: 2022-05-09.

[4] Emmanouil Alimpertis, Athina Markopoulou, Carter Butts, and Konstantinos Psounis. 2019. City-wide signal strength maps: Prediction with random forests. In *The World Wide Web Conference*. 2536–2542.

[5] Monica Alleven. 2023. AT&T dials up 5G SA uplink 2-carrier aggregation. https://www.fiercewireless.com/tech/att-dials-5g-sa-uplink-2-carrier-aggregation. Accessed: 2023-06-20.

[6] Android. [n. d.]. Android Developers: Documentation for TelephonyManager. https://developer.android.com/reference/android/telephony/TelephonyManager, accessed June 2023.

[7] T-Mobile Blog. 2024. T-Mobile Delivers Another World's First with 6-Carrier Aggregation. https://www.t-mobile.com/news/network/t-mobile-delivers-another-worlds-first-with-6-carrier-aggregation. Accessed: 2024-01-09.

[8] Jason Carpenter, Wei Ye, Feng Qian, and Zhi-Li Zhang. 2023. Multi-Modal Vehicle Data Delivery via Commercial 5G Mobile Networks: An Initial Study. In *2023 IEEE 43rd International Conference on Distributed Computing Systems Workshops (ICDCSW)*. IEEE, 157–162.

[9] Yitian Chen, Yanfei Kang, Yixiong Chen, and Zizhuo Wang. 2020. Probabilistic forecasting with temporal convolutional neural network. *Neurocomputing* 399 (2020), 491–501.

[10] Nokia Communications. 2023. 5G Carrier Aggregation explained. https://www.nokia.com/about-us/newsroom/articles/5g-carrier-aggregation-explained/#:~:text=Carrier%20Aggregation%20is%20a%20software,enhance%20the%20end%20user%20experience. Accessed: 2023-06-09.

[11] Rostand A. K. Fezeu, Claudio Fiandrino, Eman Ramadan, Jason Carpenter, Lilian Coelho deFreitas, Faaiq Bilal, Wei Ye, Joerg Widmer, Feng Qian, and Zhi-Li Zhang. 2024. Unveiling the 5G Mid-Band Landscape: From Network Deployment to Performance and Application QoE. In *Proc. ACM SIGCOMM*. Association for Computing Machinery, Sydney, NSW, Australia. https://doi.org/10.1145/3651890.3672269

[12] Rostand A. K. Fezeu, Eman Ramadan, Wei Ye, Benjamin Minneci, Jack Xie, Arvind Narayanan, Ahmad Hassan, Feng Qian, Zhi-Li Zhang, Jaideep Chandrashekar, and Myungjin Lee. 2023. An In-Depth Measurement Analysis of 5G mmWave PHY Latency and Its Impact on End-to-End Delay. In *Passive and Active Measurement*, Anna Brunstrom, Marcel Flores, and Marco Fiore (Eds.). Springer Nature Switzerland, Cham, 284–312.

[13] 5G Tools for RF Wireless. 2023. 5G NR TBS (Transport Block size) Calculator. https://5g-tools.com/5g-nr-tbs-transport-block-size-calculator/. Accessed: 2023-06-20.

[14] Moinak Ghoshal, Imran Khan, Z. Jonny Kong, Phuc Dinh, Jiayi Meng, Y. Charlie Hu, and Dimitrios Koutsonikolas. 2023. Performance of Cellular Networks on the Wheels. In *Proceedings of the 2023 ACM on Internet Measurement Conference (IMC '23)*. Association for Computing Machinery, New York, NY, USA, 678–695. https://doi.org/10.1145/3618257.3624814

[15] Signals Research Group. 2023. SRG'S SUPER BOWL SCOUTING REPORT...OF LTE AND 5G CELLULAR CAPACITY. https://signalsresearch.com/issue/srgs-super-bowl-scouting-report-of-lte-and-5g-cellular-capacity/

[16] 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.

[17] Ahmad Hassan, Arvind Narayanan, Anlan Zhang, Wei Ye, Ruiyang Zhu, Shuowei Jin, Jason Carpenter, Z. Morley Mao, Feng Qian, and Zhi-Li Zhang. 2022. Vivisecting Mobility Management in 5G Cellular Networks. In *Proc. of ACM SIGCOMM*. 86–100. https://doi.org/10.1145/3544216.3544217

[18] Ahmad Hassan, Wei Ye, Anlan Zhang, Jason Carpenter, Ruiyang Zhu, Shuowei Jin, Feng Qian, Z Morley Mao, and Zhi-Li Zhang. 2024. The Case for Boosting Mobile Application QoE via Smart Band Switching in 5G/xG Networks. In *Proceedings of the 25th International Workshop on Mobile Computing Systems and Applications*. 127–132.

[19] ITU. 2017. Minimum requirements related to technical performance for IMT-2020 radio interface(s). https://www.itu.int/pub/R-REP-M.2410-2017. Accessed: 2023-06-09.

[20] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[21] Konstantinos Kousias, Özgü Alay, Antonios Argyriou, Andra Lutu, and Michael Riegler. 2019. Estimating downlink throughput from end-user measurements in mobile broadband networks. In *2019 IEEE 20th International Symposium on" A World of Wireless, Mobile and Multimedia Networks"(WoWMoM)*. IEEE, 1–10.

[22] 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.

[23] Qianru Li, Zhehui Zhang, Yanbing Liu, Zhaowei Tan, Chunyi Peng, and Songwu Lu. 2023. CA++: Enhancing Carrier Aggregation Beyond 5G. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*. 1–14.

[24] Pingping Lin, Chunlei Hu, Xiao Li, Jinyang Yu, and Weiliang Xie. 2022. Research on Carrier Aggregation of 5G NR. In *2022 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*. IEEE, 1–5.

[25] Yuxiang Lin, Yi Gao, and Wei Dong. 2022. Bandwidth Prediction for 5G Cellular Networks. In *2022 IEEE/ACM 30th International Symposium on Quality of Service (IWQoS)*. IEEE, 1–10.

[26] Y. Liu and C. Peng. 2023. A Close Look at 5G in the Wild: Unrealized Potentials and Implications. In *Proc. of IEEE INFOCOM*. 1–10.

[27] Yanbing Liu and Chunyi Peng. 2023. A Close Look at 5G in the Wild: Unrealized Potentials and Implications. In *IEEE INFOCOM 2023-IEEE Conference on Computer Communications*. IEEE.

[28] Lifan Mei, Jinrui Gou, Yujin Cai, Houwei Cao, et al. 2022. Realtime mobile bandwidth and handoff predictions in 4G/5G networks. *Computer Networks* 204 (2022), 108736.

[29] Lifan Mei, Runchen Hu, Houwei Cao, Yong Liu, et al. 2019. Realtime mobile bandwidth prediction using lstm neural network. In *Passive and Active Measurement: 20th International Conference, PAM 2019, Puerto Varas, Chile, March 27–29, 2019, Proceedings 20*. Springer, 34–47.

[30] Dimitar Minovski, Niclas Ogren, Christer Ahlund, and Karan Mitra. 2021. Throughput prediction using machine learning in lte and 5g networks. *IEEE Transactions on Mobile Computing* (2021).

[31] Arvind Narayanan, Eman Ramadan, Jason Carpenter, Qingxu Liu, Yu Liu, Feng Qian, and Zhi-Li Zhang. 2020. A First Look at Commercial 5G Performance on Smartphones. In *Proc. of The Web Conference*. 894–905.

[32] Arvind Narayanan, Eman Ramadan, Rishabh Mehta, Xinyue Hu, Qingxu Liu, Rostand AK Fezeu, Udhaya Kumar Dayalan, Saurabh Verma, Peiqi Ji, Tao Li, Feng Qian, and Zhi-Li Zhang. 2020. Lumos5G: Mapping and predicting commercial mmWave 5G throughput. In *Proceedings of the ACM Internet Measurement Conference*. 176–193.

[33] Arvind Narayanan, Muhammad Iqbal Rochman, Ahmad Hassan, Bariq S Firmansyah, Vanlin Sathya, Monisha Ghosh, Feng Qian, and Zhi-Li Zhang. 2022. A comparative measurement study of commercial 5G mmwave deployments. In *INFOCOM 2022*. IEEE, 800–809.

[34] Arvind Narayanan, Xumiao Zhang, Ruiyang Zhu, Ahmad Hassan, Shuowei Jin, Xiao Zhu, Xiaoxuan Zhang, Denis Rybkin, Zhengxuan Yang, Zhuoqing Morley Mao, Feng Qiang, and Zhi-Li Zhang. 2021. A variegated look at 5G in the wild: performance, power, and QoE implications. In *Proc. ACM SIGCOMM*. 610–625.

[35] Corporate Communications Nokia. 2023. Nokia, Qualcomm and T-Mobile achieve industry-first 5G Carrier Aggregation combining 5 Component Carriers #MWC23. https://www.nokia.com/about-us/news/releases/2023/02/28/nokia-qualcomm-and-t-mobile-achieve-industry-first-5g-carrier-aggregation-combining-5-component-carriers-mwc23/. Accessed: 2023-06-09.

[36] Darijo Raca, Dylan Leahy, Cormac J. Sreenan, and Jason J. Quinlan. 2020. Beyond Throughput, the next Generation: A 5G Dataset with Channel and Context Metrics. In *Proc. of ACM MMSys*. 303–308. https://doi.org/10.1145/3339825.3394938

[37] Eman Ramadan, Arvind Narayanan, Udhaya Kumar Dayalan, Rostand A. K. Fezeu, Feng Qian, and Zhi-Li Zhang. 2021. Case for 5G-Aware Video Streaming Applications. In *Proc. of the 5G-MeMU*. 27–34.

[38] Muhammad Iqbal Rochman, Vanlin Sathya, Norlen Nunez, et al. 2022. A comparison study of cellular deployments in Chicago and Miami using apps on smartphones. In *Proceedings of the 15th ACM Workshop on Wireless Network Testbeds, Experimental evaluation & CHaracterization*. 61–68.

[39] Muhammad Iqbal Rochman, Vanlin Sathya, Norlen Nunez, Damian Fernandez, Monisha Ghosh, Ahmed S. Ibrahim, and William Payne. 2022. A Comparison Study of Cellular Deployments in Chicago and Miami Using Apps on Smartphones. In *Proc. of ACM WiNTECH*. 61–68. https://doi.org/10.1145/3477086.3480843

[40] Muhammad Iqbal Rochman, Wei Ye, Zhi-Li Zhang, and Monisha Ghosh. 2024. A Comprehensive Real-World Evaluation of 5G Improvements over 4G in Low-and Mid-Bands. *IEEE DySPAN'24* (2024).

[41] Alassane Samba, Yann Busnel, Alberto Blanc, Philippe Dooze, and Gwendal Simon. 2017. Instantaneous throughput prediction in cellular networks: Which information is needed?. In *2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*. IEEE, 624–627.

[42] Josef Schmid, Alfred Höss, and Björn W Schuller. 2021. A survey on client throughput prediction algorithms in wired and wireless networks. *ACM Computing Surveys (CSUR)* 54, 9 (2021), 1–33.

[43] Kevin Spiteri, Rahul Urgaonkar, and Ramesh K Sitaraman. 2020. BOLA: Near-optimal bitrate adaptation for online videos. *IEEE/ACM transactions on networking* 28, 4 (2020), 1698–1711.

[44] Sean J Taylor and Benjamin Letham. 2018. Forecasting at scale. *The American Statistician* 72, 1 (2018), 37–45.

[45] Wikipedia. 2023. 5G NR frequency bands. https://en.wikipedia.org/wiki/5G_NR_frequency_bands. Accessed: 2023-06-09.

[46] Wikipedia. 2023. LTE frequency bands. https://en.wikipedia.org/wiki/LTE_frequency_bands. Accessed: 2023-06-09.

[47] 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 *Proc. of ACM SIGCOMM*. 479–494. https://doi.org/10.1145/3387514.3405882
[48] Xinlei Yang, Hao Lin, Zhenhua Li, Feng Qian, Xingyao Li, Zhiming He, Xudong Wu, Xianlong Wang, Yunhao Liu, Zhi Liao, et al. 2022. Mobile access bandwidth in practice: Measurement, analysis, and implications. In *Proceedings of the ACM SIGCOMM 2022 Conference*. 114–128.
[49] Wei Ye, Jason Carpenter, Zejun Zhang, Rostand AK Fezeu, Feng Qian, and Zhi-Li Zhang. 2023. A Closer Look at Stand-Alone 5G Deployments from the UE Perspective. In *2023 IEEE International Mediterranean Conference on Communications and Networking (MeditCom)*. IEEE, 86–91.
[50] 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.
[51] Chaoqun Yue, Ruofan Jin, Kyoungwon Suh, Yanyuan Qin, et al. 2017. LinkForecast: Cellular link bandwidth prediction in LTE networks. *IEEE Transactions on Mobile Computing* 17, 7 (2017), 1582–1594.

APPENDIX

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

A ADDITIONAL MEASUREMENT OBSERVATIONS

Fig. 22 shows our measurement setups with different mobility patterns, and Table 5 summarizes the user equipment (and corresponding 5G modem) used in our measurements. In the following, we provide extra measurement observations in addition to **§3**.

Figure 22: Measurement setup.

Table 5: User equipment and 5G modem.

Smartphone Models (Samsung Galaxy)	S10	S20 Ultra	S21 Ultra; S21 FE	S22	S23
5G Modem Models (Qualcomm Snapdragon)	X50	X55	X60	X65	X70

A.1 Spectrum Frequency Allocation and Channel Combinations

We provide more detailed spectrum frequency allocation and channel combination information below.

Detected Channels. Table 6 reports detailed information about 4G/5G frequency channels observed during our data collection. Acknowledging that the spectrum position and bandwidth of allocated channels may exhibit slight variations across different locations (to alleviate interference), we streamline our approach by consolidating and counting them as identical channels for simplicity. We consistently have more than *four* 4G/5G channels available at each measurement location for each operator, indicating the extensive potential for carrier aggregation (CA).

Channel Combinations. Table 7 lists *select* channel/band combinations used for CA that we have observed.

Table 6: 4G & 5G channels observed in our study. 4G bands are denoted by the letter "b" and 5G by "n".

Band	Duplex Mode	Operators	Frequency (MHz)	Channel BW (MHz)	#. Ch
b2	FDD	OpX, OpY, OpZ	1,900 (mid)	5,10,15,20	18
b4	FDD	OpY, OpZ	1,700 (mid)	10,15,20	6
b5	FDD	OpY	850 (low)	10	1
b12	FDD	OpX, OpZ	700 (low)	5,10	4
b13	FDD	OpY	700 (low)	10	1
b14	FDD	OpX	700 (low)	10	1
b25	FDD	OpZ	1,900 (mid)	5	1
b29	FDD	OpX	700 (low)	5	1
b30	FDD	OpX	2,300 (mid)	5,10	2
b41	TDD	OpZ	2,500 (mid)	20	2
b46	TDD	OpX, OpY, OpZ	5,200 (mid)	20	9
b48	TDD	OpY	3,600 (mid)	10,20	17
b66	FDD	OpX, OpY, OpZ	1,700/2,100	5,10,15,20	18
b71	FDD	OpZ	600 (low)	5	5
n5	FDD	OpX, OpY	850 (low)	10	3
n25	FDD	OpZ	1,900 (mid)	20	2
n41	TDD	OpZ	2,500 (mid)	20,40,60,100	9
n66	FDD	OpX	2,100 (mid)	5,10	2
n71	FDD	OpZ	600 (low)	15,20	3
n77	TDD	OpX, OpY	3,700 (mid)	40,60,100	5
n260	TDD	OpX	39,000 (high)	100	12
n261	TDD	OpY	28,000 (high)	100	8

Table 7: Sample channel combinations observed. The last column reports the number of observed channel combinations: the first considers the ordering of SCells in the combinations and the second counts only the sets of unique channels in the combinations.

Oper.	Ch. Combinations	Aggr. BW	Num.
OpX	*4G* up to 5 CCs	Up to 90 MHz	270/162
	5G n77+n77	120 MHz	6/3
	5G up to 8 CCs in n260	Up to 800 MHz	13/3
OpY	*4G* up to 5 CCs	Up to 100 MHz	174/108
	5G n5 +n77	70 MHz	1/1
	5G n77 +n77	160 MHz	4/2
	5G up to 8 CCs in n261	Up to 800 MHz	13/8
OpZ	*4G* up to 5 CCs	Up to 90 MHz	67/42
	5G n41 + n71 + n25 + n41	Up to 180 MHz	6/2
	5G n25 + n41 + n41	160 MHz	2/1
	5G n71 + n25 + n41	135 MHz	1/1
	5G n41 + n41	140 MHz	4/3
	5G n25 + n41	120 MHz	2/1
	5G n71 + n41	up to 120 MHz	7/7

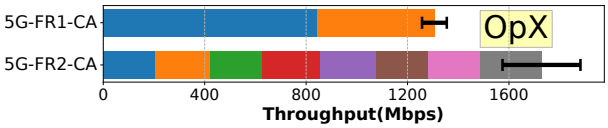


Figure 23: Throughput of OpX and OpY under the ideal channel condition.

A.2 Impact of CA on 4G/5G Performance.

We now present the impact of CA on 5G (and 4G) performance for all three surveyed mobile operators under various mobility and scenario settings.

Ideal Channel Condition. Fig. 1 and Fig. 23 report the achievable throughput for all three mobile operators under *stationary* mobility with the line-of-sight to the base station. For the downlink, 4G and 5G mmWave (FR2) throughput increases exponentially due to each SCell using a bandwidth similar to that of PCell. In the 5G

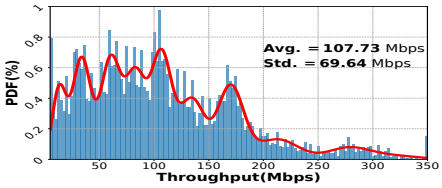


Figure 24: 4G throughput distribution.

low-/mid-band (FR1), OpZ also achieves an average of 1.5 Gbps by integrating three smaller bandwidth channels as SCells, resulting in more than double the throughput of not using CA. Even though a lower number of SCells are configured, OpX and OpY still achieve an average of 1.3 Gbps and 1.6 Gbps throughput with the aggregation of C-band. Despite news [5] of its deployment, we only observe CA on the uplink for 5G mmWave, which has 321.5 Mbps uplink throughput in limited locations.

Outdoor Driving. Fig. 25 shows the prevalence of CA observed from our *driving* experiments. We see that CA is widely used for 4G services by all three carriers, whether it is urban, suburban, or along highways. As for 5G services, OpZ has deployed CA aggressively, not only in urban environments but also in around 75% of the suburban areas surveyed and along with many areas along highways. Meanwhile, OpX and OpY's 5G CA deployment has also achieved significant progress, covering around 25% and 54% of the surveyed urban areas, where their high-band aggregation takes up 6% and 25%, respectively. For a comparison of the 4G and 5G throughput using the same driving measurement data, we refer to Fig. 26. We see that CA boosts the 4G throughput for all mobile operators to nearly, and occasionally well above, 100 Mbps in the urban and suburban areas, as well as along the highways. Surprisingly, with CA in FR1 channels, OpZ can nonetheless boost its *average* 5G throughput to more than 700 Mbps, 600 Mbps, and 350 Mbps in the urban, suburban, and highway settings, respectively. While less CA is observed while *driving*, urban locations for OpX and OpY have approximately double the throughput when compared to their suburban locations at nearly 450 Mbps and 840 Mbps, respectively.

Indoor Walking. We have also conducted indoor *walking* experiments, which create a more challenging channel environment. From Fig. 27, we see that for all three mobile operators, the indoor downlink throughput experiences a significant drop compared to the results obtained under ideal channel conditions. Meanwhile, OpX and OpY have a high probability of dropping back to 4G. In contrast, OpZ uses the low-band FDD channel as PCell (which extends coverage due to it experiencing less radio path loss and thereby receives higher power, as shown in Fig. 28) and another mid-band TDD channel as SCell (which increases bandwidth) to form FDD-TDD CA. OpZ can expand its indoor 5G coverage and achieve good throughput performance.

Impact of UE Capability. CA not only depends on the availability of channel combinations, but also on the capability of UE. This is particularly true when the number of 5G CCs increases. To account for this, we conduct experiments using three Samsung smartphone models. Fig. 29 shows the throughput and percentage of CC detected on the walking dataset for Samsung S10, S21 and S22 phones.

Impact of CC Changes. During our driving experiments in urban, suburban, and highway scenarios, we observed that the addition and removal of CCs occurs on average every 33.97s, 28.83s, and 16.11s [6], respectively. These events result in average throughput changes of 176%, 248%, and 1016% within a 5-second window, resulting in the standard deviation of 212 Mbps, 89 Mbps, and 140 Mbps. In contrast, when only considering the impact of other factors with a stable connection, the throughput standard deviation is 123 Mbps, 50 Mbps, and 70 Mbps.

Table 8: The signal strength (in dBm) of each CC (run on the same channel and connected to the same PCI) is measured at different times of the day at the same location. T1 represents peak hours, while T2 and T3 correspond to non-peak hours.

	CC-1 ($n41^a$)	CC-2 ($n41^b$)	CC-3 ($n25$)	CC-4 ($n71$)
T1	-99.8 ± 3.3	$-100.4.8 \pm 1.9$	-99.2 ± 1.5	-89.2 ± 1.4
T2	-99.5 ± 3.1	-102.2 ± 3.1	-98.5 ± 1.4	-88.1 ± 1.7
T3	-96.7 ± 2.1	-99.1 ± 2.0	-97.3 ± 1.8	-89.4 ± 1.5

B THROUGHPUT FACTORS

B.1 Theoretical Channel Capacity

Each transport block size (TBS) can be approximated as

$$\begin{aligned} TBS &= Quantizer(N_{\text{info}}) \\ &= Quantizer(N_{\text{re}} \cdot R \cdot Q_m \cdot v) \end{aligned} \tag{1}$$

where N_{re} is the number of resource element. R denotes the coding rate, the ratio between the number of information bits and the total number of bits sent, and Q_m denotes the maximum modulation order (e.g., 6 for 64QAM and 8 for 256QAM). Lastly, v is the MIMO layers, representing the spatial multiplexing. They altogether decide the number of information bits, i.e., N_{info}. The $Quantizer$ will further consider PHY layer processing, such as encoding, and add round bits to fit the system. See [13] for more details.

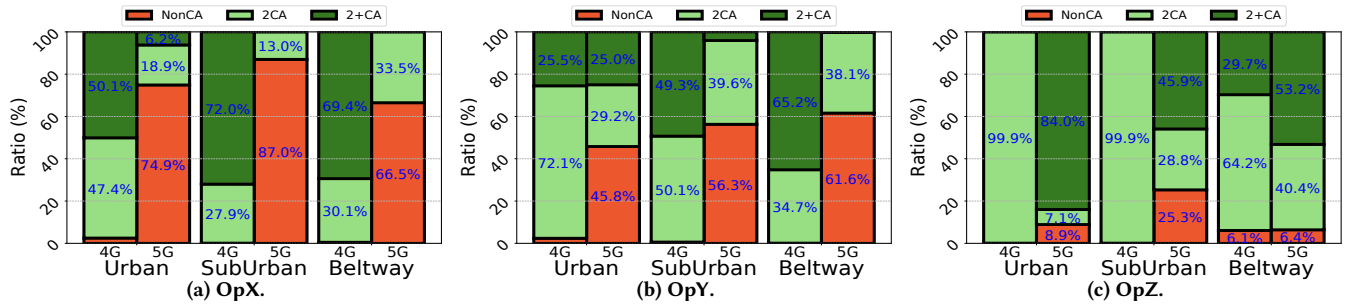
B.2 Temporal Dynamics.

Although the key measurement results presented in this paper are primarily collected during the cell low-traffic periods to avoid traffic throttling. We also conduct measurements at different times of the day to show the impact of temporal dynamics. We find that the cell's overall performance remains stable over weeks or months since operators typically do not frequently change hardware or configurations. When examining temporal dynamics on a daily scale, we observe that the signal strength of each carrier component (connecting to the same PCI band) converges, as shown in Table 8.

However, we also observe user numbers will cause temporal dynamics and impact throughput. To investigate this issue, we picked two locations on campus with good and bad signal coverage. For simplicity, we visualize one component carrier's throughput traces in Fig. 31 and Fig. 32. The traces are collected at different times of the day, including the rush hour when thousands of students move among buildings between two classes (T1) and non-rush hours (T2 and T3) The color represents the throughput, with green indicating higher throughput and red indicating lower throughput.

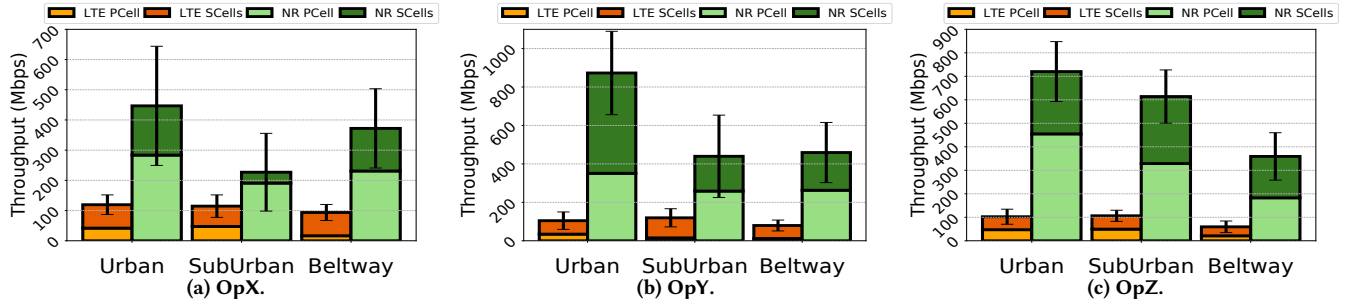
Our data reveals that UE will experience significantly lower throughput during rush hour (T1), particularly in locations with suboptimal signal strength. This is because the cell may allocate

[6] Highways drive faster, while urban areas need to stop for traffic lights.



(a) OpX. (b) OpY. (c) OpZ.

Figure 25: CA occurrence during outdoor driving experiments.



(a) OpX. (b) OpY. (c) OpZ.

Figure 26: Throughput performance of the outdoor driving experiment.

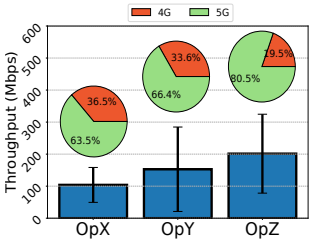
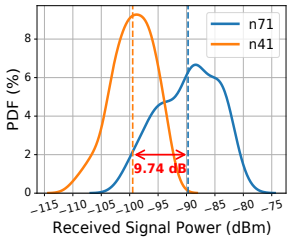
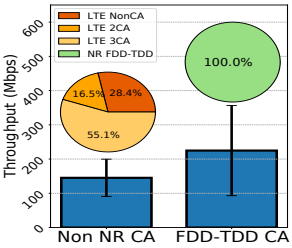


Figure 27: Indoor walk throughput. The top pie chart reports the technology occurrence rate.



(a) Low-band (n71) receives higher signal power under challenge scenarios.



(b) Lock/unlock the use of FDD low-band (n71).

Figure 28: FDD-TDD CA expands 5G coverage.

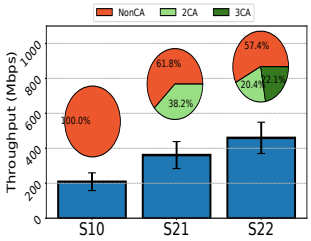
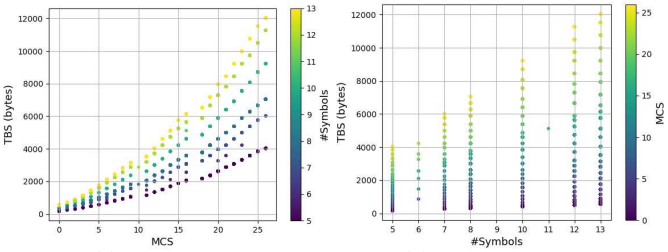
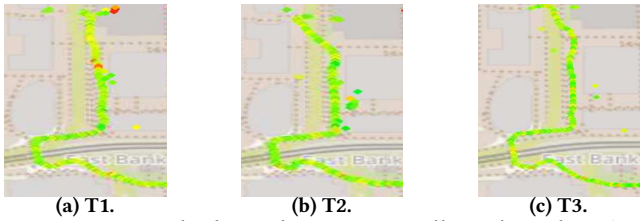


Figure 29: S10 does not support SA-5G CA, S21 supports 2CC CA and S22 supports up to 3CC CA.



(a) TBS vs MCS. (b) TBS vs #symbols.

Figure 30: Side view of Fig. 9.



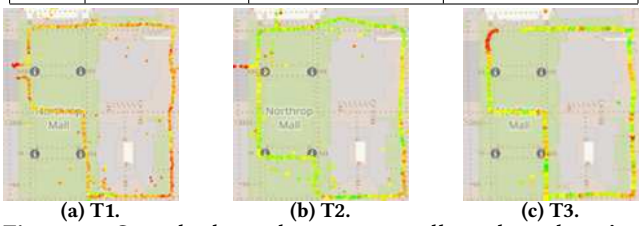
(a) T1. (b) T2. (c) T3.

Figure 31: Sample throughput traces collected at a location with *good* coverage at different times of the day. The green color indicates good throughput.

fewer resource elements/blocks. For detailed statistics, see Table 9 and Table 10. On the other hand, the other PHY layer parameters, such as channel conditions (CQI), modulation and coding rate (MCS), show little variance. This validates that the temporal dynamic of 5G throughput can be captured by those features, which makes our modeling feasible. In the CA cases, the total number of aggregated resource blocks decreases at the peak hour, while the signal strength and channel quality (CQI) of each cell remain consistent. Interestingly, we also find that PCell and SCell occasionally decrease at varying degrees, and SCell may be dropped. We believe this is more related to RAN-side scheduling algorithms, and will leave it for future studies.

Table 9: Statistic values of cases in Fig. 31.

	T1: rush hour	T2: non-rush hour	T3: non-rush hour
CQI	10.0 ± 2.5	10.9 ± 2.0	10.5 ± 2.1
MCS	**16.4 ± 4.0**	17.5 ± 3.2	17.3 ± 3.2
#RB	**61.0 ± 19.3**	70.8 ± 21.6	66.1 ± 18.2



(a) T1. (b) T2. (c) T3.

Figure 32: Sample throughput traces collected at a location with *bad* coverage at different times of the day. The green color indicates good throughput.

Table 10: Statistic values of cases in Fig. 32.

	T1: rush hour	T2: non-rush hour	T3: non-rush hour
CQI	6.2 ± 1.5	6.1 ± 1.4	6.3 ± 1.4
MCS	**11.1 ± 2.8**	11.2 ± 2.7	11.2 ± 2.7
#RB	**64.4 ± 22.9**	88.1 ± 22.1	95.0 ± 32.6

Table 11: Statistic of the dataset used for ML.

Idx	Oper.	Tech	Mobility	Scenarios	Size
1	OpX	4G/5G	walk	Outdr-urban, Indr	Each scenario contains 10 traces with 300 to 600 data samples per trace at a time scale of 10ms and 1s.
2	OpX	4G/5G	driving	Urban, Suburban, Beltway	
3	OpY	4G/5G	walk	Outdr-urban, Indr	
4	OpY	4G/5G	driving	Urban, Suburban, Beltway	
5	OpZ	4G/5G	walk	Outdr-urban, Indr	
6	OpZ	4G/5G	driving	Urban, Suburban, Beltway	

C ADDITIONAL ML IMPLEMENTATION AND EVALUATION RESULTS.

In this section, we provide additional information about the machine learning experiment setups and results discussed in **§6**.

C.1 Experiment Setups

Datasets. Table 11 summarizes the dataset used for machine learning. Table 12 describes the data field.

Implementation. We implement *Prism5G* using PyTorch framework. Each RNN and MLP module has a two-layer architecture with 128 hidden units. The input sequence length is set to 10, while the output sequence length is 10. We normalize the dataset using the min-max scaler and randomly divide the dataset into training, validation, and test sets using a ratio of 0.5:0.2:0.3 for all experiments. *Prism5G* and all the deep learning-based baselines are trained using the Adam optimizer [20] with a learning rate of 0.01, a batch size of 128, and a max epoch of 200. We utilize root-mean-square-error (RMSE) as the loss function. The best ML model is selected according to its performance on the validation dataset. All data processing and numerical experiments are run on a workstation, which runs on Ubuntu 20.04 system with a 64-Core AMD Ryzen Threadripper PRO 3995WX, 1TB main memory, and 3× Nvidia RTX A6000 GPUs.

Table 12: Illustration of data fields.

Features	Illustrations	Type
5G Modem	The 5G modem chipset	Binary
Event	Signaling control events for CC activation and deactivation	Binary
Band Info	Band and channel index	One-hot
ssRSRP	Synchronize signal reference signal received power [dBm]	Float
ssRSRQ	Synchronize signal reference signal received quality [dB]	Float
SINR	Signal to interference & noise ratio [dB]	Float
CQI	Channel quality indicator	Float
BLER	Block level error rate [%]	Float
#RB	Number of allocated resource blocks	Float
#Layer	Number of MIMO layers	Float
MCS	Modulation and Coding Schema	Float
HisTput	Historical throughput [Mbps]	Float

Train. & Eval. strategy for Prophet. The Prophet algorithm is assessed using the cross-validation schema[7]. It will continuously update the training set by incorporating new historical data samples as the time-domain sliding window moves. This allows us to refit the Prophet algorithm for prediction. This methodology grants Prophet an advantage over neural network-based models, which solely rely on a fixed training set. However, it also comes with increased computation time.

Train. & Eval. strategy for classical ML. The GBDT (gradient-boosted decision trees) and RF (random forest) only take the feature for the prediction and don't have an intrinsic design for the time series task. We combine all historical data into a single feature (i.e., $R^{(T,k)}$-> $R^{(T\times k,1)}$) and feed it into the algorithm as a whole. We choose the regression tree as our aim is to predict the throughput.

Ablation study of *Prism5G*. Various components of *Prism5G* were removed for an ablation study to demonstrate their necessity. The performance of these simplified models in RSME compared to the full model are shown in Table 13.

C.2 Additional Results

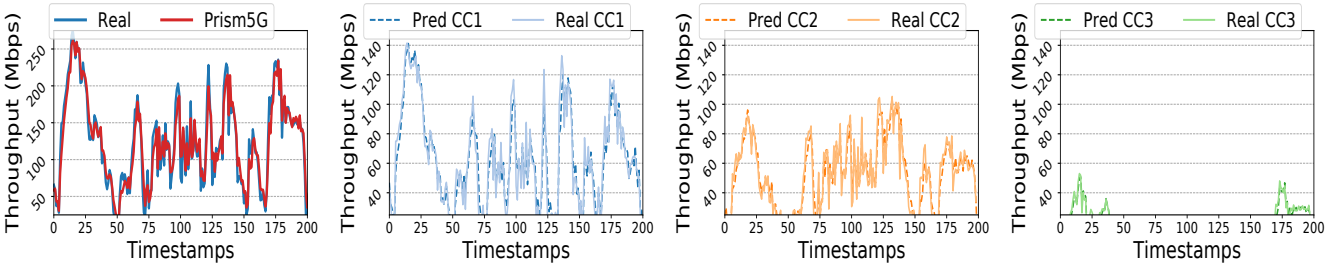
Performance of baselines at transition points. Fig. 33 and Fig. 34 display the results of inference performed by *Prism5G* on the urban downtown driving dataset for 4G and 5G compared to real observed values. A comparison of the performance of Prophet and the LSTM against the real values to show their difficulty in modeling around transition points is shown in Fig. 35 and Fig. 36, respectively.

Generlizability. Table 14 presents the results of evaluations conducted on (1) the same route but different runs as the training dataset and (2) entirely new routes not included in the training

[7] https://facebook.github.io/prophet/docs/diagnostics.html

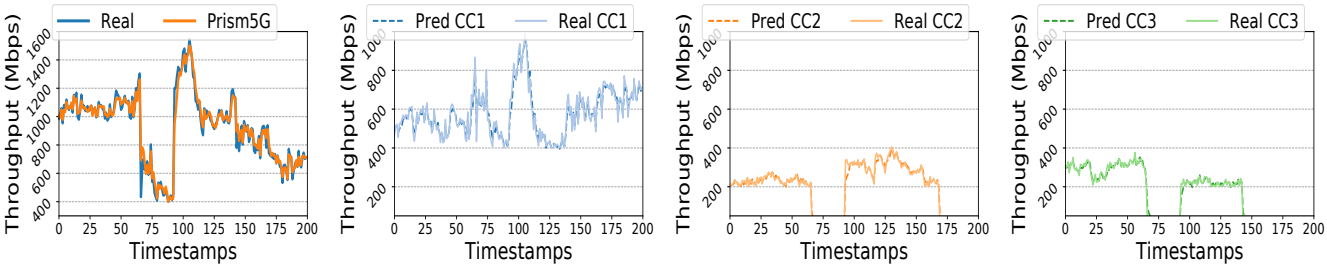
Table 13: Performance results of the ablation study of *Prism5G* measured in RMSE, lower values being better.

Dataset	Short			Long		
	No State	No Fusion	Prism5G	No State	No Fusion	Prism5G
OpX (Walking)	0.195	0.201	**0.188**	0.198	0.202	**0.187**
OpX (Driving)	0.289	0.298	**0.283**	0.213	0.221	**0.200**
OpY (Walking)	0.207	0.211	**0.195**	0.205	0.210	**0.192**
OpY (Driving)	0.292	0.287	**0.274**	0.280	0.279	**0.260**
OpZ (Walking)	0.255	0.262	**0.240**	0.245	0.234	**0.228**
OpZ (Driving)	0.374	0.370	**0.352**	0.282	0.283	**0.277**



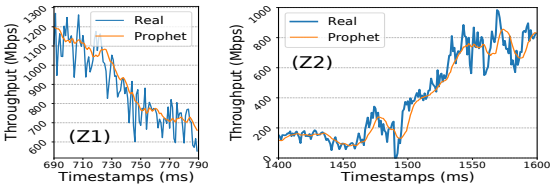
(a) Aggregated 4G Tput. (b) PCell. (c) SCell[1]. (d) SCell[2].

Figure 33: Visualized *Prism5G* Prediction Results for 4G on urban downtown driving data.



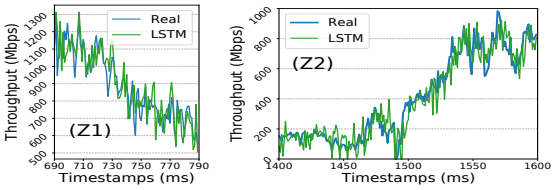
(a) Aggregated 5G Tput. (b) PCell. (c) SCell[1]. (d) SCell[2].

Figure 34: Visualized *Prism5G* Prediction Results for 5G on urban downtown driving data.



(a) Throughput drop. (b) Throughput boost.

Figure 35: Prophet performance at the transition point area (Z1 and Z2) with a *10-ms* time scale.



(a) Throughput drop. (b) Throughput boost.

Figure 36: LSTM performance at the transition point area (Z1 and Z2) with a *10-ms* time scale.

dataset. These tables serve as a means to assess the generalizability of the proposed *Prism5G*.

Table 14: Evaluation of Prism5G generalizability on (1) the same route but different runs and (2) the new routes not included in the training dataset. We use RMSE as a metric, with lower values being better.

	Prophet	LSTM	Lumos5G	Ours	Improv. (%)
(1)	0.478	0.330	0.361	0.299	9.4%
(2)	0.512	0.379	0.352	0.308	12.5%