

# PUBLICATIONS

Home / Publications / ComSoc Technology News (CTN) / Revisiting Cloud RAN from a Computer Architecture Point of View

## Revisiting Cloud RAN from a Computer Architecture Point of View

IEEE CTN

**Written By:** Q. Zheng, Y. Chen, S. Abeyratne, R. Dreslinski, and T. Mudge, The University of Michigan, Ann Arbor  
 Published: 1 Jul 2016

**STAY IN THE KNOW!**  
[JOIN THE CTN MAILING LIST](#)



CTN Issue: July 2016

**A note from the editor:**

This month we have asked Prof. Trevor Mudge's team to give us an update on some of the work they are doing in cloudRAN L1 architecture. Trevor's team has a best in class reputation for computer architecture generally and comes at this problem from that view point. In particular Qi's thesis [1] is well worth a read. We hope you enjoy their summary below. Comments always welcome.

Alan Gatherer, *Editor-in-Chief*

### Designing General Purpose Cloud Platforms for Future Radio Access Networks

Q. Zheng, Y. Chen, S. Abeyratne, R. Dreslinski, and T. Mudge, The University of Michigan, Ann Arbor

The number of mobile device users has increased rapidly over the last decade. Only 3 years ago there were 335 million wireless subscribers and 300 thousand base stations in the United States, making wireless communication a market worth \$200 billion annually. A crucial component of wireless communication is a radio access network (RAN), which connects mobile devices and the core network. Because of the need for 24/7 service availability and the growing requirements for high data rate, RAN systems consume significant energy and capital. Already in 2010, wireless base stations consumed 110 million kWh of energy and cost \$40 billion on capital expenditure. This constrains the traditional RAN growth in terms of both energy consumption and total cost of ownership (TCO). In addition, the throughput of traditional RANs cannot meet the growing demand for higher data rates. Global mobile traffic increased 66-fold with a compound annual growth rate of 131% from 2008 to 2013, while the peak throughput of the wireless network only increased at 55% annually. This has resulted in reduced data rates per user. For example, the typical user download speed of LTE is only 10% of the specification's peak data rate. The throughput of the traditional RAN is not adequate and will only worsen with new applications such as 4K videos demanding high data rates. Consequently, finding solutions to improve throughput, energy, and cost of RAN systems is highly desirable.

To solve the problems that constrain the traditional RANs a new emerging cloud service has been proposed, cloud radio access network (C-RAN). C-RAN is a domain specific cloud service that combines the traditional RAN with cloud computing technology. In C-RAN, the non-compute intensive remote radio heads (RRHs) are decoupled from the compute intensive baseband units (BBUs): RRHs remain at the distributed base station sites while BBUs are aggregated into a centralized cloud datacenter. All distributed sites are connected to the datacenter through a high speed front-haul link (see Fig. 1). Prototype C-RAN systems have been deployed by China Mobile, SK Telecom, and Korea Telecom. DoCoMo in Japan has recently announced that they will deploy C-RAN in their LTE-A networks.

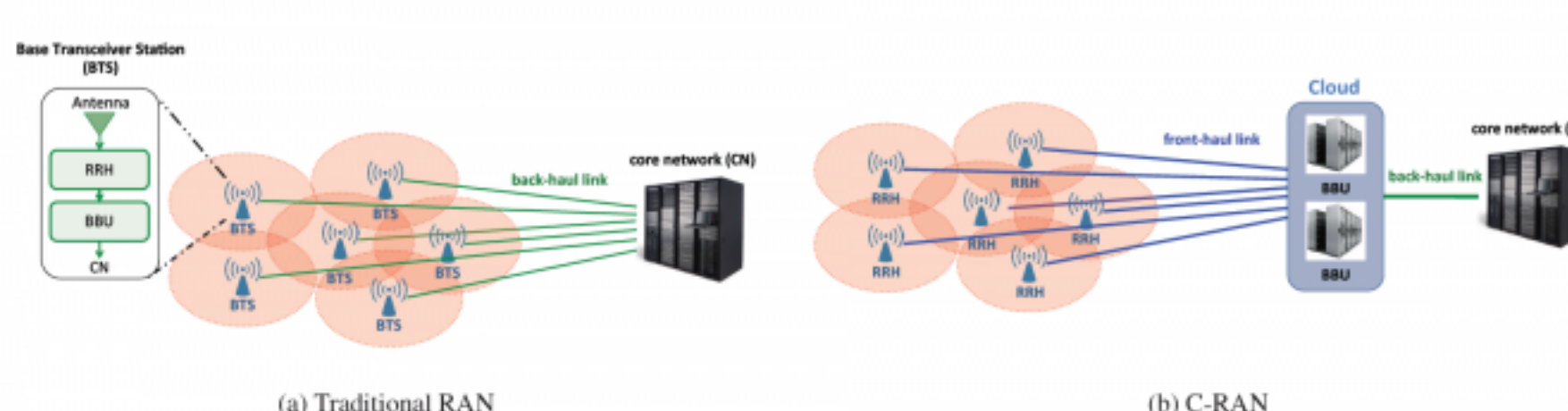


Figure 1: The structures of a traditional RAN system and a C-RAN system. In a traditional RAN, each BTS has an RRH and a BBU, and BTSs are distributed to cover each site. In C-RAN, only RRHs are distributed, and the BBUs sit in a centralized location, i.e. the cloud.

C-RANs have many advantages including reduction in energy and TCO and improvement in throughput and hardware utilization. On the front-end, removing the BBU from the base station makes them smaller and simpler, which reduces the energy and the TCO of the site. For example, site acquisition and rental fees are smaller as are electricity costs and hardware upgrade costs. In addition, because the sites are smaller, more of them can be deployed in densely populated areas, which improves the quality of service. On the back-end, aggregating BBUs into a centralized datacenter saves maintenance cost, and improves hardware utilization and energy efficiency by sharing computing resources among sites. It also increases network capacity by enabling joint processing (a technique to reduce interference from multiple base stations when a mobile device is at the edge of a coverage area). Higher hardware utilization, lower energy, and lower cost, would allow operators to deploy more hardware to improve the throughput. Although C-RANs have been proposed and deployed for future wireless systems, there remain open design questions about high-speed links, fast I/O and datacenter design for this new technology. Our work (see Q. Zheng's thesis [1]) was focused on the C-RAN datacenter design. There are several datacenter design questions that they answered, including: **What is the best general purpose platform for C-RAN datacenters? What is the most power-efficient and cost-efficient design? How can C-RANs be designed to handle future growth?**

To resolve these key design concerns, we believe a well-designed CRAN datacenter should achieve the following targets:

1. Meet the throughput requirement specified in current and future wireless standards with commodity servers.
2. Minimize the energy consumption and the TCO.
3. Manage hardware resources to handle the temporal and spatial imbalances in traffic.
4. Support the number of sites required by the current CRAN design, and be able to scale up for larger C-RANs in the future.

To understand these challenges we created a model of the C-RAN BBU uplink receiver that includes the key kernels in the physical (PHY) layer and the Turbo decoder [2]. The focus was on the receiver rather than the transmitter as it has most of the computations in the C-RAN BBU. Next, we investigated how this model performs on commodity general purpose servers in two major platforms, which are multi-core CPUs and GPUs. We have implemented the LTE BBU model in both C++ and CUDA for the evaluation on CPUs and GPUs, respectively. For the C++ implementation, we maximize performance by using automatic vectorization and openMP optimizations. For the CUDA implementation, we explored various types of parallelism to maximize performance [3].

In our evaluation we compared CPU servers and GPU servers across performance, energy, and TCO. For the performance, we compared the throughput achieved by each type of server to the throughput defined by the LTE specification and determined the amount of equipment needed to be deployed in a C-RAN datacenter supporting 32 sites. We found that the GPU servers consistently achieve better performance than the CPU servers. This is because the data and thread level parallelism present in many of the BBU kernels are better suited for the GPU architecture. Our results show that we need 4x to 16x as many CPU servers as the GPU servers in the equivalent datacenter (Fig. 2). For 32 sites, the CPU-based datacenter consumes on average 13x more energy (Fig. 3) and has 6x higher TCO (Fig. 4) than the GPU-based datacenter.

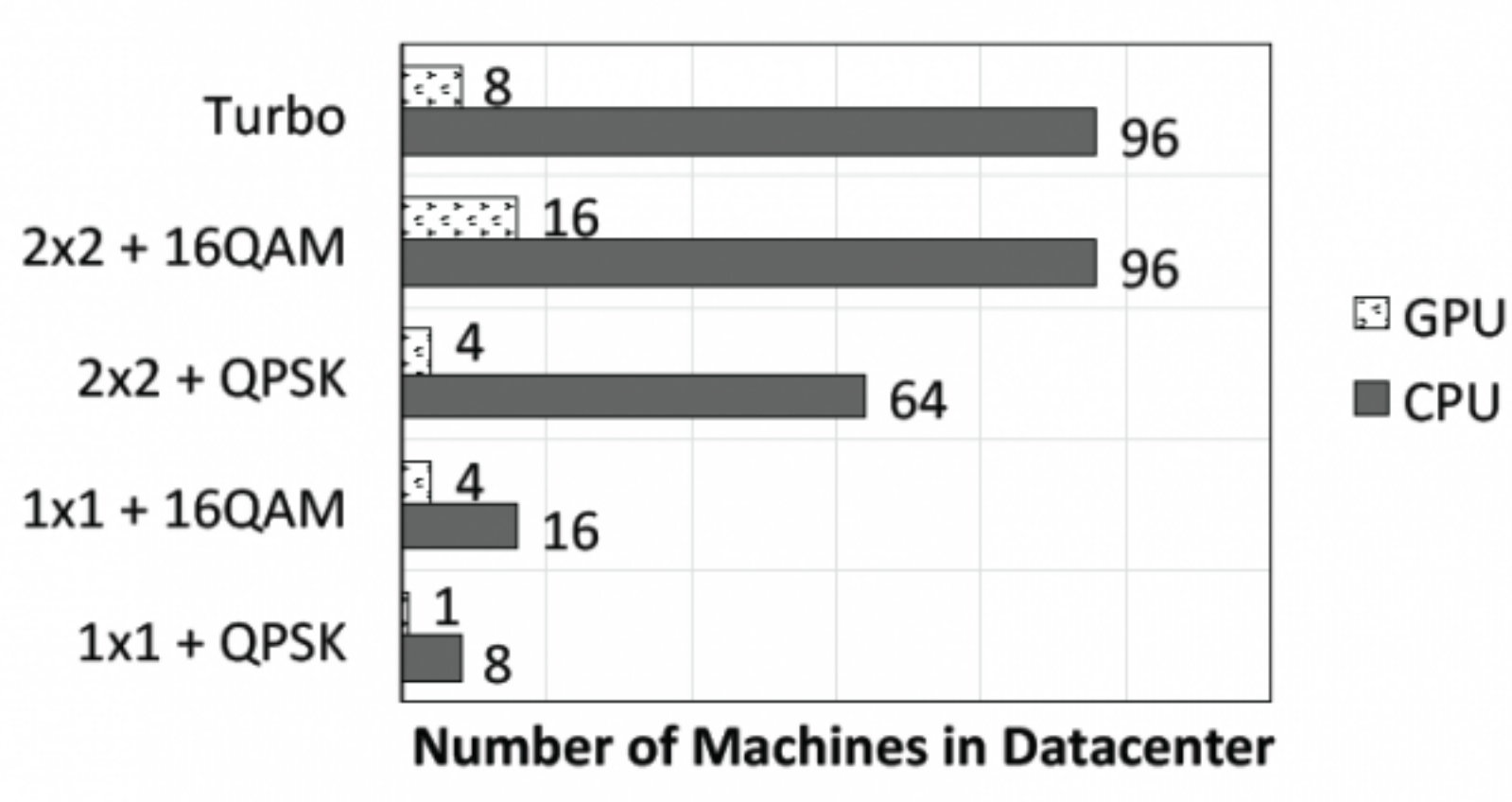


Figure 2: Number of servers to realize a 32-site datacenter

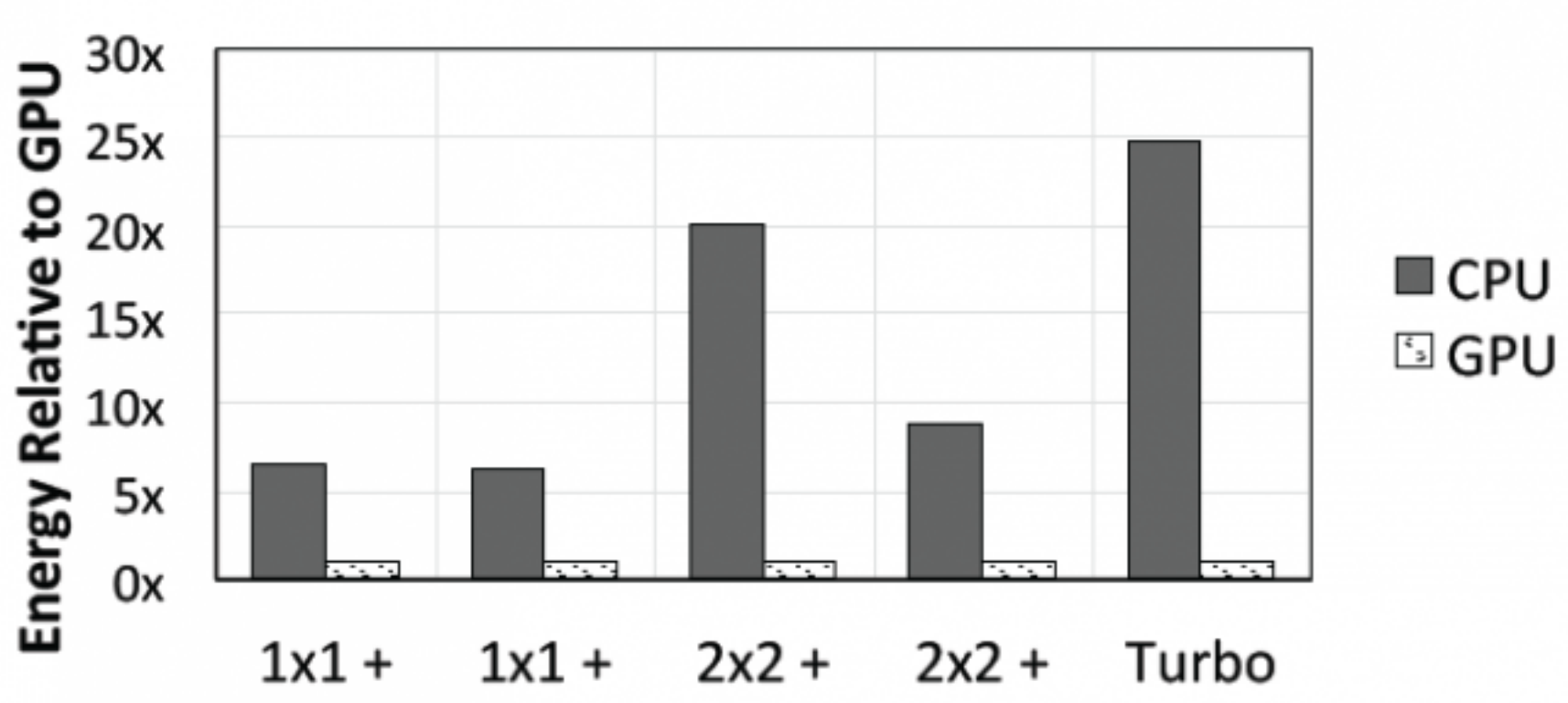


Figure 3: Energy required to process 32 sites at full load.

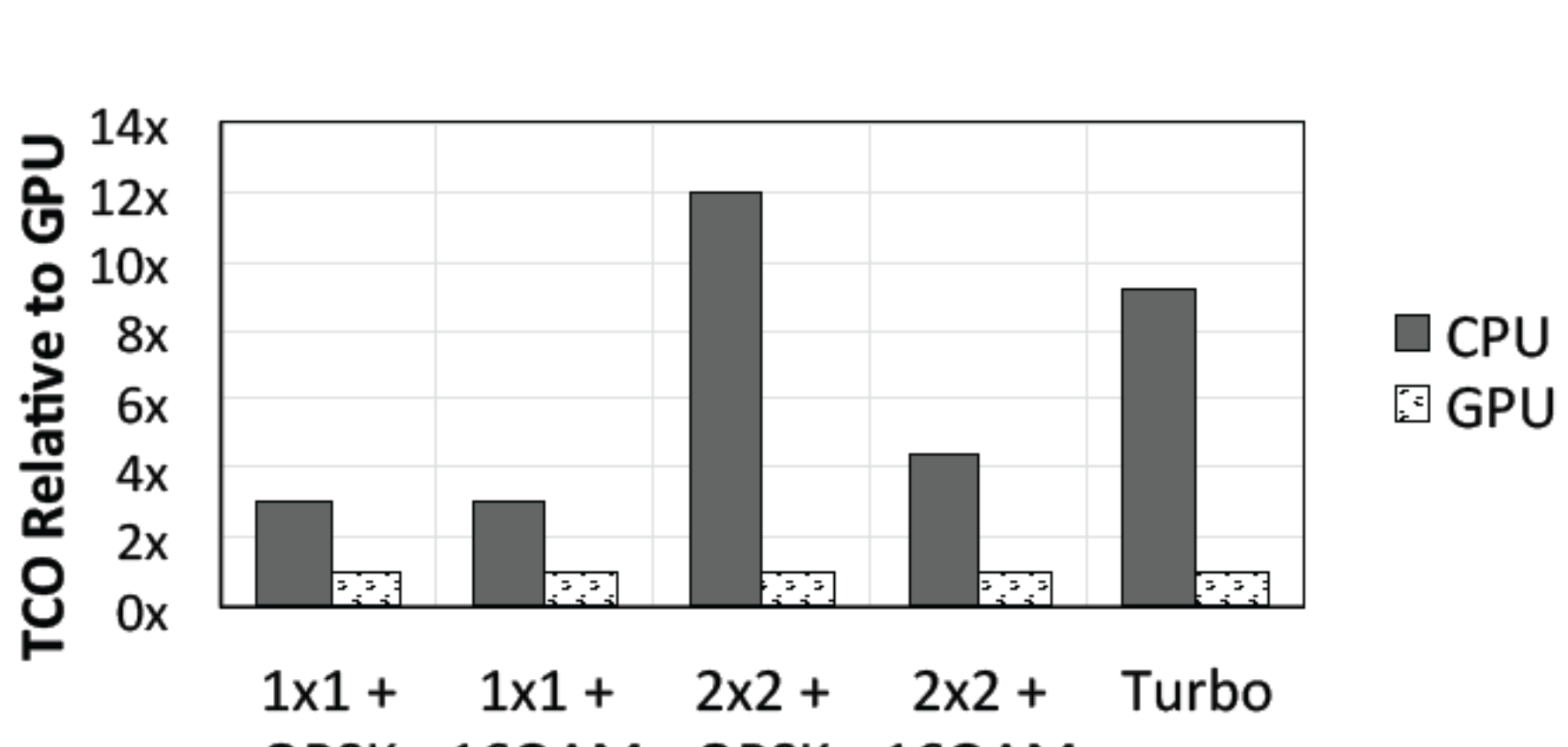


Figure 4: TCO of the 32-site datacenter.

### References

1. Q. Zheng. Datacenter Design for Future Cloud Radio Access Network. Ph.D. Thesis. The University of Michigan, 2016.
2. Q. Zheng, Y. Chen, R. Dreslinski, C. Chakrabarti, A. Anastasopoulos, S. Mahlke, and T. Mudge, "WiBench: An open source kernel suite for benchmarking wireless systems," in 2013 IEEE International Symposium on Workload Characterization (IISWC), pp. 123–132, Sept 2013.
3. Q. Zheng, Y. Chen, H. Lee, R. Dreslinski, C. Chakrabarti, A. Anastasopoulos, S. Mahlke, and T. Mudge, "Using Graphics Processing Units in an LTE Base Station," Journal of Signal Processing Systems, vol. 78, no. 1, pp. 35–47, 2015.

**Statements and opinions given in a work published by the IEEE or the IEEE Communications Society are the expressions of the author(s). Responsibility for the content of published articles rests upon the authors(s), not IEEE nor the IEEE Communications Society.**

[SIGN IN TO COMMENT](#)

CONNECT WITH US: