Key Considerations for vRAN:

Insights from SK Telecom and NTT DOCOMO

February 2024

SK TELECOM & NTT DOCOMO VRAN WHITE PAPER

Table of Contents

List of abbreviations

6G: 6th Generation mobile communication systems

AAL: Acceleration Abstraction Layer

AI: Artificial Intelligence

API: Application Programming Interface

CAPEX: Capital Expenditure

CNF: Container Network Function

COTS: Commercial Off The Shelf

CPU: Central Processing Unit

C-RAN: Centralized RAN

CSI: Channel State Information

CU: Central Un

D-RAN: Distributed RAN

DU: Distributed Unit

FEC: Forward Error Correction

gNB: next generation Node B

IT: Information Technology

KPI: Key Performance Indicator

L1: Layer 1

LCM: Life Cycle Management

MAC: Media Access Control

MEC: Multi-access Edge Computing

MIMO: Multiple-Input and Multiple-Output

ML: Machine Learning

MNO: Mobile Network Operator

OAM: Operation, Administration, Maintenance

OPEX: Operating Expenditure

PCIe: Peripheral Component Interconnect express

PDCP: Packet Data Convergence Protocol

PHY: Physical Layer

RAN: Radio Access Network

RF: Radio Frequency

RLC: Radio Link Control

RRC: Radio Resource Control

RU: Radio Unit

SDAP: Service Date Adaptation Protocol

SoC: System on a Chip

TCO: Total Cost of Ownership

TDD: Time Division Duplex

UE: User Equipment

VNF: Virtual Network Function

vRAN: virtualized RAN

1 Executive summary

This white paper provides key considerations for the successful vRAN deployment and operation from the perspective of MNOs. It includes an analysis of vRAN architectures with a focus on L1 accelerator, which is a key component for successful vRAN deployment. It also highlights the importance of selecting an appropriate accelerator type that aligns with the MNOs' network architecture and requirements associated with their deployment scenarios.

The introduction of vRAN brings various benefits, such as flexibility in software and hardware combinations, enhanced RAN maintenance and operation through virtualization, and infrastructure sharing from the edge to the core network. For this reason, vRAN is expected to play a significant role in the 6G era. However, ongoing research and development in vRAN technologies are necessary to address existing challenges, such as further enhancing cell capacity and power efficiency compared to traditional RAN. Therefore, it is critical to clarify key considerations for successful vRAN deployment from the perspective of MNOs, while considering the evolution of future networks.

Furthermore, this white paper presents an analysis of vRAN architectures with primary focus on L1 accelerators. vRAN is designed to run on COTS server hardware, which is manufactured as general purpose-hardware, and L1 acceleration technologies play a crucial role in vRAN operation. Currently, L1 Acceleration technology can be classified into two types: (i) CPU-integrated look-aside L1 accelerators, which have evolved from PCIe-integrated look-aside L1 accelerators, and (ii) Inline L1 accelerators. Some manufacturers advocate for the benefits of a look-aside structure, while others assert that an inline structure is a better solution for successful vRAN deployment. However, the advantages and disadvantages of each L1 accelerator type can be interpreted differently depending on the MNO's network architecture and requirements. Therefore, it is essential for MNOs to make a carefully considered and well-informed selection of L1 accelerators, taking into account their network architecture and policies. This white paper provides insights through an analysis on cell capacity, energy efficiency and complexity from the perspective of MNOs.

2 Introduction

2.1 Overview of vRAN

There is a growing demand for high processing performance in base station equipment to accommodate the increasing data rates and capacity in mobile communications networks. To meet this demand, MNOs have been utilizing base station equipment that is implemented with purpose-built hardware solutions.

On the other hand, the IT industry has experienced significant technological advancements in virtualization and cloud computing. These advancements have made it possible to separate hardware from software, together with improved hardware performance. As a result, MNOs are now exploring the use of virtualization technology for their base station equipment. This involves the utilization of COTS servers, which are general-purpose hardware components, along with hardware accelerators specifically designed for processing radio signals (referred to as L1 accelerators). This new approach enables the realization of a vRAN.

Typically, base station equipment consists of three components: CU, DU, and RU. In vRAN, CU and DU components are virtualized (VNF) or containerized (CNF), and these virtualized CU (vCU) and DU (vDU) are deployed on COTS servers, as shown in Figure 1.

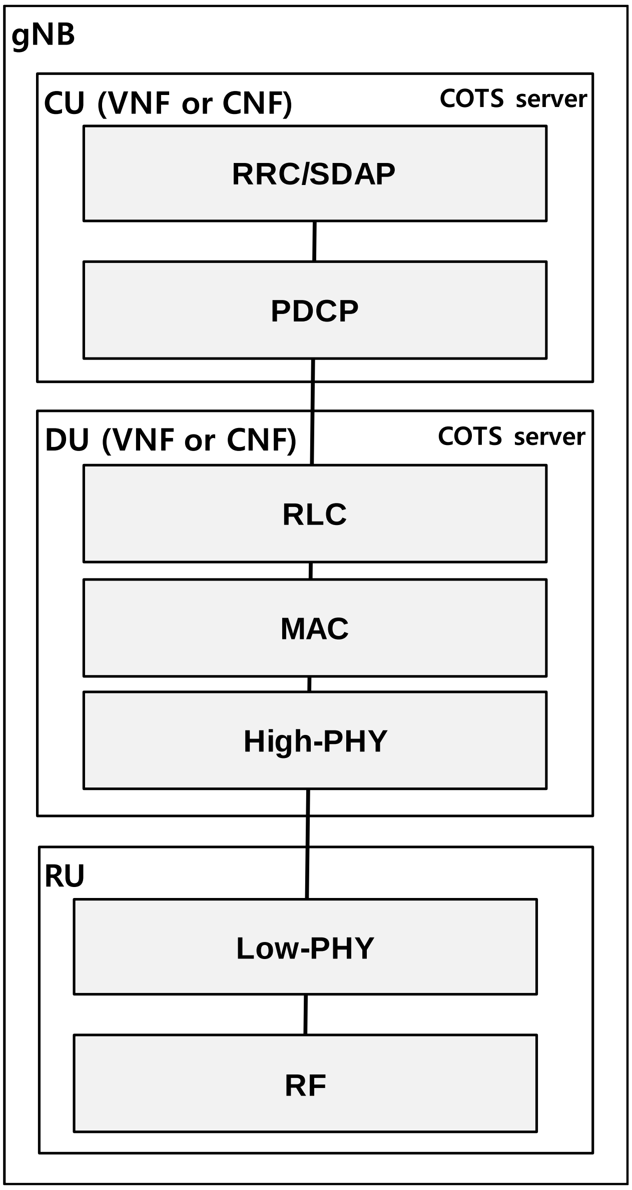

Figure1: Base station for vRAN

Benefits of introducing vRAN

A key feature of vRAN is the decoupling of software from hardware in implementation of CU/DU using the virtualization technology. This feature is expected to bring the following benefits:

1. Flexibility in software and hardware combinations:
    - Inclusion of new software vendors can be considered, which enables avoiding vendor lock-in as well as collaboration with specialized software vendors and/or open-source communities.
    - Leveraging competition in the existing ecosystem for COTS servers can lead to cost reduction of base station equipment. Additionally, utilizing COTS servers with the latest hardware technologies can improve performance in virtualized CU/DU.

2. Enhanced RAN maintenance and operation through virtualization:
    - The virtualized platform for CU/DU, deployed on COTS servers, can be centrally managed by software, resulting in improved RAN maintenance and operation.
    - Virtualization technology enables automated healing and scaling, allowing for quick expansion and reduction of virtualized CU/DU within the pooled resources. This enables shortened lead times and increased flexibility (e.g., network slicing) in vRAN deployment.

3. Infrastructure sharing and common operations from edge to core network:
    - The same virtualized platform can be used from the edge network, such as for vRAN and Multi-Access Edge Computing (MEC), to the core network. This facilitates efficient maintenance and operation of the entire networks through unified life cycle management (LCM) of virtual resources and applications.

Challenges to be addressed in vRAN

RAN represents most of the infrastructure investment for MNOs. However, the virtualization of RAN has not progressed as much as that of the core network. To facilitate the deployment of vRAN, the following aspects needs to be addressed.

Firstly, realizing a cost-effective vRAN in terms of both CAPEX and OPEX is essential for MNOs, which achieves reduced TCO for deploying the entire vRAN system.

Secondly, achieving high performance in vRAN is crucial, which is enabled by successful integration of base station software, virtualization platforms, and COTS servers, including L1 accelerators. In particular, the integration of L1 accelerators for massive physical layer processing and real-time signal processing with COTS servers is necessary to handle the demanding processing requirements of RAN.

2.2 Scope and purpose of this white paper

The scope and purpose of this white paper is to provide technical key considerations of successful vRAN deployments and an analysis of L1 accelerators from the perspective of MNOs.

While various vendors are developing vRAN solutions, there is still room for improvement to fully realize anticipated benefits of vRAN. Therefore, this white paper provides key considerations from the perspective of MNOs that are essential for vRAN realization.

Especially, L1 accelerators play a crucial role in offloading the CPU load for L1 processing and are indispensable for successful vRAN deployment. Therefore, it is important for MNOs to understand the suitable types of L1 accelerators for their vRAN deployments. Although there are some existing vendor-led white papers that analyze L1 accelerators based on their considerations (such as [1], [2], and [3]), this white paper is independently developed by SK Telecom and NTT DOCOMO. It presents the pros and cons of different L1 accelerator types from the viewpoint of MNOs.

3 Key Considerations for vRAN

This chapter focuses on the key technologies and evolution directions for vRAN realization from the perspective of MNOs. The objective is to highlight the critical aspects that are essential for the successful deployment and operation of vRAN.

The evolution of the L1 accelerator is essential for enhancing performance of cell capacity and power consumption in accordance with network architecture.

While traditional RAN software is operated on purpose-built hardware, vRAN software is designed to run on COTS server hardware, which is manufactured as general purpose-hardware. In general, vRAN implementation using general-purpose hardware still have room to further enhance the performance. To achieve this, L1 acceleration technologies play a crucial role.

From the perspective of MNOs, achieving a comparable level of cell capacity and energy efficiency to traditional RAN are key requirements of vRAN. The L1 accelerator is a key technology for meeting these MNO requirements. However, the requirements of vRAN are not one-dimensional. They can vary based on network architecture and each MNO's specific requirements such as operational policies. Therefore, it is important to understand which L1 accelerator type is more suitable for each MNO's network deployment. As a result, this white paper provides L1 accelerator analysis in the following chapter.

Key features for vRAN such as resource pooling, scaling and auto-healing should be prioritized.

Currently, major vendors primarily focus on stabilizing vRAN performance (e.g., improving capacity and power consumption) and providing functionalities equivalent to those already implemented in traditional RAN using purpose-build hardware. Consequently, the development of distinct features unique to vRAN has been postponed for future consideration. As a result, full potential of vRAN has not been fully realized yet despite its anticipated benefits.

One example of a vRAN specialized feature is resource pooling between DUs. Leveraging virtualization technologies, resources can be shared and dynamically allocated among vDUs. If a particular DU experiences high traffic load, resources can be reallocated from another DU to that DU. vRAN can also optimize the resources of each server using features such as scale in/out. In addition, in the case of vRAN, the functions are virtualized and can be recovered without any service interruption. This is much faster and more effective compared to managing individual hardware directly, minimizing business downtime and maintaining productivity.

Therefore, to facilitate widespread adoption of vRAN, it is important to develop specialized features and use cases that provide unique benefits achievable with vRAN.

Energy efficiency is an essential KPI for MNOs to moving towards vRAN.

One of the main key requirements in adopting vRAN is to reduce its power consumption. Improvements in energy efficiency can be achieved not only through advancements in hardware technologies, including those in L1 accelerators, but also through a new development and implementation of vRAN-specific power-saving software features.

For instance, energy efficiency can be enhanced by effectively controlling the Active/Inactive states of chipsets, like the CPU, based on the load conditions of the vRAN. Optimizing vRAN software while considering the characteristics of general-purpose hardware can also help to reduce processing power and energy consumption. Additionally, leveraging virtualization benefits like resource pooling and reallocation among DUs can contribute to overall energy savings across a mobile network.

The enhancement of management technologies, tools, and procedures plays a crucial role in improving the integration of vRAN.

Since vRAN consists of various hardware and software components from diverse vendors, integration is one of the key requirements for adopting vRAN. Therefore, orchestration function is essential to automate and centrally manage the configuration, provisioning, and scaling of vRAN components. In addition, advanced monitoring and analytics tools are essential for managing vRAN deployments. These tools should provide real-time visibility of the performance and health of vRAN equipment, through monitoring and analytics of fault, configuration, and performance managements. Furthermore, effective LCM of virtualized functions is essential for smooth integration of vRAN.

By leveraging advanced management technologies, tools, and procedures, MNOs can improve the integration of vRAN deployments, enhance the operational efficiency, optimize the performance, and consequently deliver high-quality services to end-users.

TCO, network controllability, and standardization should be considered for future network based on cloud and AI native.

As cloud and AI/ML technologies gain more popularity, various new approaches in RAN architecture are considered.

The use of cloud-native technology is expected to further increase in vRAN due to its advantages, such as flexible deployment. Cloud-native technology also enables easy expansion of new functionalities, like AI/ML.

There is ongoing progress in utilizing AI/ML technology throughout a mobile network, including RAN. Firstly, AI-based RAN automation performs both AI learning and inference on the OAM (Operation, Administration, Maintenance) server, based on statistics extracted from base stations, to improve operating efficiency. Secondly, AI-based RAN operation aims to maximize the desired KPIs by leveraging data, such as internal information of the base station, UE reports, and adjacent base station information. Lastly, AI-based air interface applies AI to the L1 interface between the base station and UE, including functions such as Channel State Information (CSI) compression, beam management, and improved positioning accuracy [4].

Since these approaches are still in the preliminary stage, it is unclear how architectural evolution will take place in 6G era and which technologies will be centered in the future. While conducting research on technical evolution, MNOs need to consider several factors. Firstly, it is important to ensure that the introduction of new architectural changes does not increase TCO, and in particular, securing computing resources for cloudification is an issue that should be carefully and realistically considered. Also, as AI-native and cloud-native architectures make network-generated information accessible to various vendors and solution providers, MNOs need to maintain control and privacy of network-related information. Lastly, in order to support interoperability between various functions, it is important to standardize APIs for both MNOs and venders [5].

Security of vRAN equipment is crucial.

Similar to traditional RAN, MNOs prioritize ensuring security in vRAN deployments. According to [6], the virtualization platform in vRAN equipment is associated with a wider range of threats compared to other components and interfaces. These threats pertain to hardware resources, virtual machines/containers, and virtualization layer, and they can have an impact on other components of vRAN as well as secure network communication. Therefore, it is essential to address these threats adequately and enhance the security of vRAN equipment. In this regard, any security-related activities pertaining to vRAN developments (e.g., [7]) hold great significance for MNOs.

4 vRAN Architecture using L1 Acceleration Technology

In a traditional RAN, a single network vendor usually provides purpose-built hardware and corresponding software to the market. Meanwhile, in vRAN, where a key advantage is the separation of hardware and software for implementing CU and DU functions, it is possible to use different vendors' solutions, such as COTS hardware, software for virtualization platform and that for RAN functions in vCU/vDU.

However, a major concern when implementing vRAN on the COTS hardware is securing the necessary computing capacity for the L1 processing, which includes FEC and modulation functions. There is a fear that the workload of High-PHY layer may exceed the capability of a general-purpose CPU. Consequently, the use of L1 accelerators is being considered as a solution to either partially or fully offload the workload of CPUs for High-PHY layer processing. In simple words, L1 accelerators are essential for the successful implementation of vDU. Therefore, it is important for MNOs to well understand the characteristics of L1 acceleration technologies, taking into account their network architecture and requirements. This chapter aims to provide analysis and insights on vRAN architecture focused on L1 acceleration from the perspective of MNOs.

4.1 Technology & market status

Initially, selected-function L1 accelerators, also known as look-aside accelerators, were adopted in the form of an integrated PCIe card. The look-aside accelerator handles specific functions in the High-PHY layer, such as FEC, while the CPU handles the remaining High-PHY layer functions. Currently, the L1 accelerators' architecture is evolving in two directions: one is the CPU-integrated look-aside L1 accelerators and the other is the Inline L1 accelerators, as shown in Figure 2. The CPU-integrated look-aside L1 accelerators integrate the look-aside accelerators and CPU in a SoC, which removes the need for separate PCIe cards. On the other hand, unlike the look-aside approach, the Inline accelerator is designed to fully cover all High-PHY layer functions.

While a separate card is required for the Inline accelerators in COTS servers, they offer the advantage of reducing CPU load more than the look-aside accelerators do.

At present, the market status is at a stage where the two types of acceleration technology are competing. From the perspective of MNOs, it is important to understand which type of accelerator is more appropriate for their network architecture. Therefore, conducting an in-depth analysis of the pros and cons of each accelerator type is necessary before deploying vDU equipment in their networks.

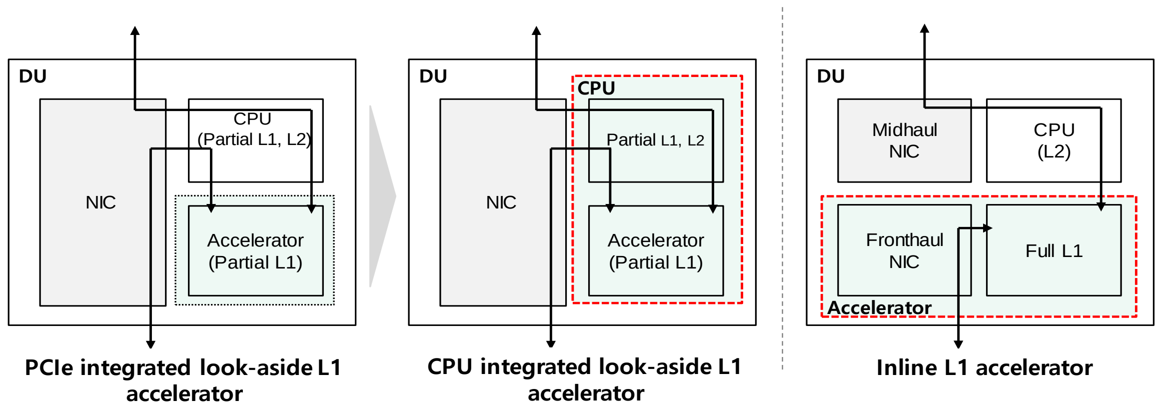

Figure 2. Different type of L1 accelerators

4.2 Analysis of acceleration technologies

As explained previously, acceleration technology is presently evolving from the PCIe-integrated look-aside L1 accelerators to the CPU-integrated look-aside L1 accelerators or the Inline L1 accelerators. Some manufacturers contend the benefits of a look-aside structure, while others assert that an Inline structure is a better option for deploying vRAN.

However, the requirements for deployment of vRAN are not one-dimensional. They can vary based on network architecture and specific requirements of each MNO, such as cell capacity, energy efficiency, and operational policies. Therefore, it is important for MNOs to understand which accelerator type is more suitable for their network architectures, based on a thorough analysis of the advantages and disadvantages of each type. To analyze the pros and cons of these different types of L1 accelerators, we considered two types of RAN architectures, namely C-RAN and D-RAN, as shown in Figure 3.

D-RAN is a network architecture where the DU is located at each cell site. One advantage of D-RAN is that it relaxes fronthaul requirements because the distance between the DU and the cell site is shorter than in C-RAN.

On the other hand, C-RAN is a network architecture where multiple DUs are centralized in a single location, while RUs are deployed at respective cell sites. In the C-RAN architecture, network management and maintenance can be centralized, and inter-DU resource pooling can be efficiently executed thanks to the centralization of DUs.

From the perspective of MNOs, it is important to analyze energy efficiency, cell capacity, complexity, etc. of each accelerator type based on their network architecture when they consider deployment of vRAN in their networks.

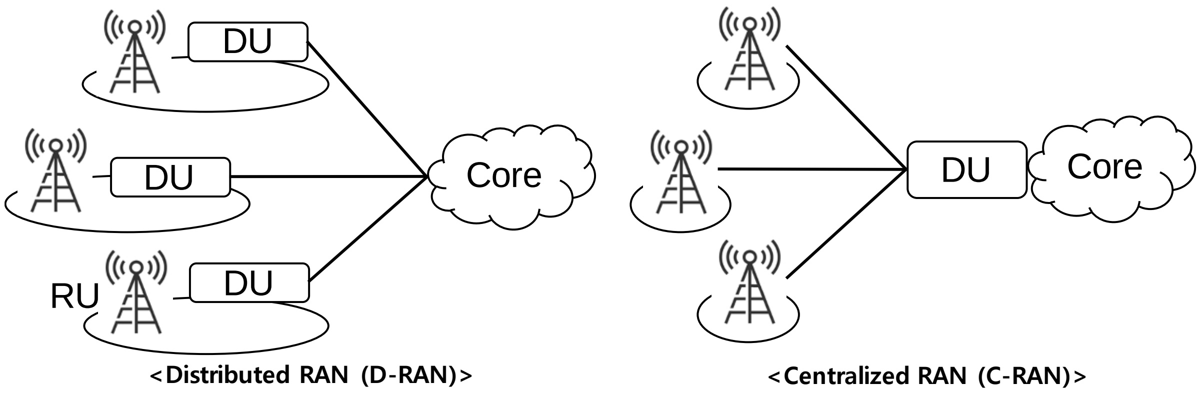

Figure 3. Deployment of D-RAN and C-RAN

Cell capacity & scalability

The number of cells that DUs can accommodate directly affects the investment cost by MNOs. This capacity is a critical requirement for the successful commercialization of vRAN.

In the case of the CPU-integrated look-aside L1 accelerators, the number of accelerators is contingent on the number of CPUs per server. Typically, a single CPU chipset is installed in a COTS server, and this limits the scalability of the CPU-integrated look-aside L1 accelerators because it is hard to increase the maximum number of cells that a single CPU can handle. On the other hand, the inline L1 accelerators can reduce CPU workload by offloading all High-PHY processing, and multiple inline L1 accelerators can be installed in a single COTS server. These advantages make more flexible and scalable vRAN deployments for MNOs.

Generally, the inline L1 accelerators would be the preferred option in the C-RAN, which can support for a large number of cells. Meanwhile, in the D-RAN architecture, the cell capacity and scalability could be a less significant factor in deciding whether to employ the look-aside or inline L1 accelerators.

However, the requisite cell capacity can vary based on each MNO's requirements and operational policies, such as per-cell traffic load, number of MIMO layers per cell, CPU specifications, and software implementation by vendors. Therefore, MNOs need to access and decide their preferred acceleration architecture and requirements through a detail analysis about the cell capacity.

Power consumption & energy efficiency

Power consumption is one of major key factors for MNOs. When there are various vCU/vDU equipment from different vendors and they provide similar capacity in terms of the supported bandwidth and performance, the power consumption of vCU/vDU could be a deciding factor for procurement.

Figure 4 provides an analysis of "Total power consumption" and "Power consumption per cell" for the look-aside L1 accelerators. The "Total power consumption" refers to the overall power consumption of the vDU, including the accelerators, while the "Power consumption per cell" represents the value derived by dividing the "Total power consumption" by the total number of supported cells. For this analysis, we assumed there are 4 MIMO layers in both downlink and uplink, a 100 MHz TDD bandwidth, and a 100% traffic load. It is important to note that power consumption may vary depending on the given conditions.

Compared to the PCIe-based look-aside L1 accelerator, the CPU-integrated look-aside L1 accelerator consumes less power because it does not require a separate accelerator for specific High-PHY functions. Moreover, the higher cell capacity of the CPU-integrated look-aside L1 accelerator compared to the PCIe-based look-aside L1 accelerator leads to improved power consumption per cell.

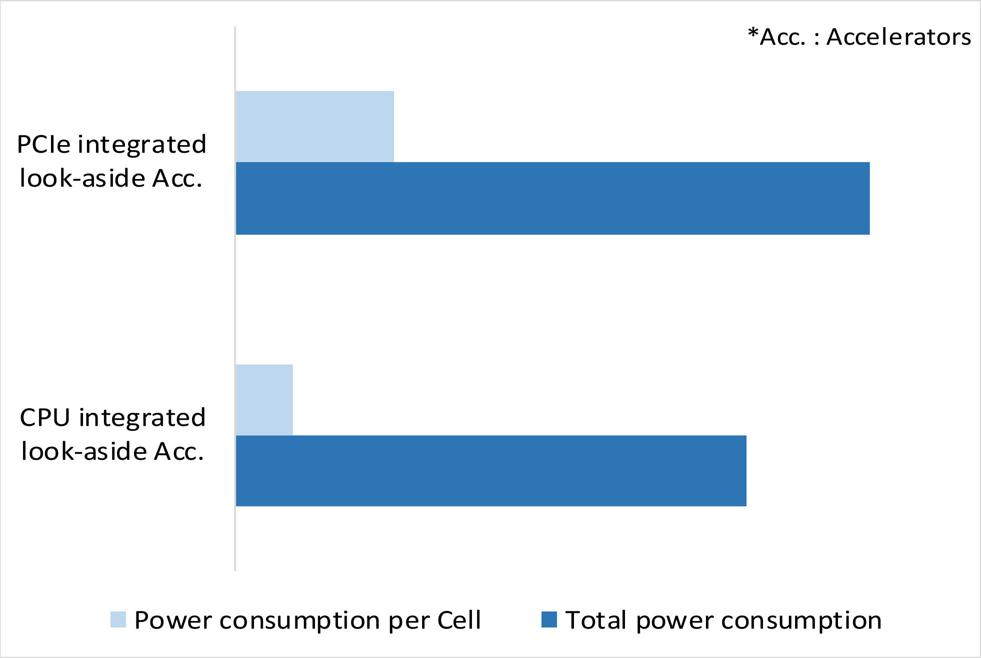

Figure 4. Power consumption of look-aside L1 accelerator

Similar to the previous figure, Figure 5 presents an analysis of "Total power consumption" and "Power consumption per cell" for the inline L1 accelerators. In the case of inline L1 accelerators, additional accelerator cards can be installed in the vDU to meet MNO’s requirements, and we considered 1 to 3 inline L1 accelerator cards per COTS server. As the number of accelerator cards increases, the total power consumption increases. However, this increase in accelerator cards results in an increase in cell capacity, and consequently, leading to a decrease in power consumption per cell.

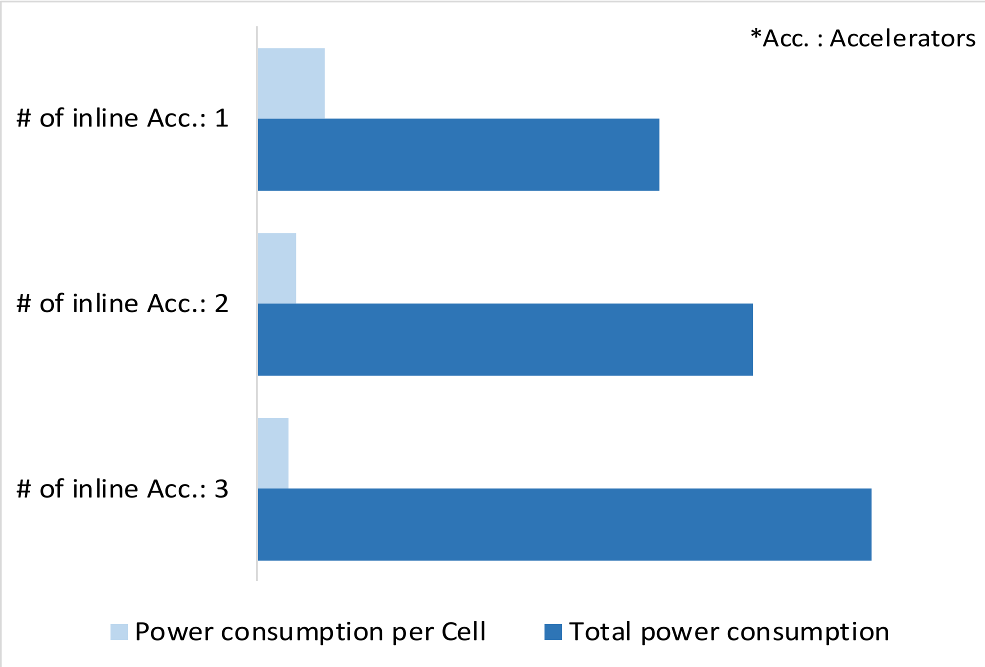

Figure 5. Power consumption of Inline L1 accelerators

The integration of accelerator and CPU functions in the CPU-integrated look-aside L1 accelerator would result in lower total power consumption compared to the inline accelerator, which requires separate hardware installation. Hence, the CPU-integrated look-aside L1 accelerators can be considered a suitable solution in the D-RAN architecture where high cell capacity would not be a major requirement, but low power consumption per site is important. Meanwhile, the C-RAN architecture requiring high cell capacity would benefit from the use of inline L1 accelerators. They allow the power consumption per cell to decrease when multiple accelerator cards are added for high cell capacity.

It is important to note that power consumption can vary depending on the given test conditions, such as traffic load per cell, MIMO layers per cell, and other factors. Furthermore, the specifications of CPUs and L1 accelerators, as well as software implementations by vendors, can also affect the results. Therefore, MNOs should carefully evaluate power consumption considering their specific network conditions and requirements.

Design & operation complexity versus flexibility

vDU has a decoupled hardware and software structure, which can offer MNOs the flexibility of potentially installing any vendor's DU software onto specific COTS hardware. However, it is important for MNOs to note that design complexity also becomes a pivotal factor in DU virtualization for ensuring smooth interoperability.

Inline L1 accelerators handle all High-PHY processing using their own hardware and software solutions, and MNOs can independently select an accelerator vendor based on their requirements, separate from the DU software vendor. While the inline L1 accelerators offer such flexibility, there could be increased complexity in ensuring interoperability between the High-PHY and other layers if vDU software and the inline accelerator manufacturers are different. Integration and maintenance complexity could also be increased since MNOs have to additionally consider an accelerator vendor for integration in the inline L1 accelerators in vDU. Moreover, MNOs could experience additional integration difficulties when they attempt to change vendors for the inline L1 accelerators.

On the other hand, the CPU-integrated look-aside L1 accelerator processes selected functions of High-PHY layer, such as FEC, while the CPU handles the majority of High-PHY layer functions. This could simplify the integration of vDU software vendors as they only need to consider certain APIs for the selected High-PHY functions.

Considering the above, while inline L1 accelerators provide flexibility, in terms of structure, it is expected that the CPU-integrated look-aside L1 accelerators have the advantage of ensuring less complexity for the vDU compared to the inline L1 accelerators. It should be noted that, to address the complexity issue for the inline accelerators, O-RAN ALLIANCE is now developing the specification regarding Acceleration Abstraction Layer (AAL) [8]. This specification enables decoupling of RAN software from hardware by defining abstract interfaces for accelerated RAN functions to allow a RAN software implementation to work with different accelerator implementations. This standardization activity is crucial for MNOs to allow for flexibility of selecting L1 accelerator cards and to avoid vendor lock-in.

5 Conclusion

vRAN is gradually spreading in MNOs' networks with its potential to provide various benefits, such as flexibility in software and hardware combinations and is also expected to be applied in the 6G era. However, existing challenges to further enhance performance and to fully achieve anticipated benefits necessitate ongoing research and development in vRAN technologies.

For the wider adoption of vRAN, it is crucial that these technological advancements evolve in the right direction, accommodating the requirements of MNOs. In this context, key considerations such as the evolution of L1 acceleration technology, virtualization-specific features, power-saving technologies, integration improvement, and security aspects should be taken into account for the vRAN technology evolution.

Among these considerations, the analysis of vRAN architectures with a focus on L1 acceleration is important. vRAN is designed to run on COTS server hardware, which is manufactured as general purpose-hardware and L1 accelerators play an essential role. From the perspective of MNOs, the advantages and disadvantages of each L1 accelerator type can be interpreted differently depending on the network architecture and specific requirements. Therefore, it is important for MNOs to well understand the characteristics of L1 acceleration technologies and select an appropriate L1 accelerator type considering their network architecture and requirements. Furthermore, additional evolution of accelerator and research on new structures should be conducted to satisfy the diverse requirements of various MNOs.

References

[1] NOKIA, “Cloud RAN: A Guide to Acceleration Options”, https://onestore.nokia.com/asset/213050?_ga=2.175314210.1675305126.1705849268-895889297.1691563930, 2023.

[2] Ericsson and Verizon, “Cloud RAN Acceleration Technology”, https://www.ericsson.com/4ae403/assets/local/ran/doc/cloud-ran-acceleration-technology-positioning-paper.pdf, 2022.

[3] Fujitsu, “What is the difference between inline and lookaside accelerators in virtualized distributed units?”, https://www.fujitsu.com/global/documents/products/network/Whitepaper-Inline-vs-Lookaside-Virtualized-vCUDU.pdf, 2023.

[4] SK Telecom, “SK Telecom 6G White paper”, https://www.sktelecom.com/en/press/press_detail.do?idx=1575, 2023.

[5] SKT and NTT DOCOMO, “6G Common Requirements”, https://www.docomo.ne.jp/english/binary/pdf/corporate/technology/rd/docomo6g/6GRequirementsWhitePaper_22February2023.pdf, 2023.

[6] Quad Critical and Emerging Technology Working Group, “Open RAN Security Report”, https://www.ntia.gov/report/2023/open-ran-security-report, May 2023.

[7] O-RAN ALLIANCE website: https://www.o-ran.org/blog/the-o-ran-alliance-security-working-group-continues-to-advance-o-ran-security.

[8] O-RAN Working Group 6, “O-RAN Acceleration Abstraction Layer General Aspects and Principles 7.0”, O-RAN.WG6.AAL-GAnP-R003-v07.00 Technical Specification, Oct. 2023.