White Paper

intel.

Open vRAN Radio Intelligent Controller Expands RAN Capability

MAVENIR™

Executive Summary

Open vRAN deployments are built with disaggregated and virtualized software components, connected through open and standardized interfaces, and interoperable with different vendors. Open vRAN systems permit building of multivendor networks leading to lower total cost of operation (TCO) and higher innovation opportunities.

Network operators are increasingly looking into building intelligence into their networks to maximize their return on investment and minimize operational costs. This is also the goal of the O-RAN Alliance[1] which has developed a standard for the RAN intelligent controller (RIC). Using the innovative RIC concept, AI/ML enabled solutions on near-Real Time RIC (near-RT RIC) and non-Real Time RIC (non-RT RIC) platforms will optimize network energy savings, and various user-level and cell-level performance metrics.

Various use cases can be supported via applications running on the RIC platform i.e., xAPPs and rAPPs on the near-RT RIC and the non-RT RIC respectively. These include use cases such as:

- Traffic steering that allows operators to optimize the network and UE performance in order to achieve balanced cell load by distributing UEs between cells to optimally achieve throughput or low delays.

- Energy savings optimization enabled by AI/ML algorithms on the RIC that identify opportunities for cell shutdown to maximize aggregate energy efficiency KPIs across all the cells.

This white paper examines optimization of cloud-native Open RAN systems using the RAN Intelligent Controller. The focus is on the implementation aspects of cloud-native virtualized Open RAN systems powered by Intel® architecture-based commercial servers and containerized and highly optimized RAN software developed by Mavenir and Intel.

1. Introduction

5G cellular networks provide improved wireless communications and networking capabilities, enabling heterogeneous use cases such as Ultra Reliable and Low Latency Communications (URLLC), Enhanced Mobile Broadband (eMBB), and massive machine-type communications (mMTC). The networks must permit a range of deployment scenarios to support a range of use case requirements.

Open RAN is gaining momentum as an alternative to the traditional single-vendor closed RAN. Open interfaces and software-defined protocol stacks enable the integration of intelligent, data-driven closed-loop control for the RAN. The O-RAN Alliance is a global community of mobile network operators, vendors, and research and academic institutions coming together to develop Open RAN standards and interoperability. The Open RAN (O-RAN) paradigm, standardized by the O-RAN Alliance and established on the foundations of software-defined networking and network function virtualization, is based on the following principles:

▪ **Open Interfaces:** The interfaces between different RAN functional nodes in O-RAN are open for multi-vendor interoperability. Open RAN enables operators to use equipment from different vendors, while ensuring interoperability between them. Open and standardized interfaces also allow operators to onboard different equipment vendors, which opens the RAN ecosystem to smaller players.

▪ **Virtualization:** The network functions are migrated from vendor-proprietary nodes to virtualized loads on commercial-off-the-shelf (COTS) hardware. Disaggregation and virtualization, based on cloud-native principles, enable flexible deployments. This increases the resiliency and reconfigurability of the RAN. With its virtualization capabilities, an Open vRAN system allows the servers to be shut down or be in various power saving state modes during off-peak times or be dedicated to other applications to reduce idle mode power consumption.

▪ **Intelligence:** The RAN is open to radio resource management (RRM) through third-party applications deployed in centralized RICs. The RICs host applications performing closed-loop control of the RAN over open interfaces, leveraging advanced data-driven AI/ML techniques to efficiently learn complex cross-layer interactions across nodes and optimize RRM decisions not captured by traditional RRM heuristics.

▪ **Programmability:** The network is programmatically configured and adapted with AI/ML-driven policies, based on continuous monitoring of network and user equipment (UE) performance.

The O-RAN Alliance has defined the Open RAN disaggregated architecture on top of 3GPP LTE and new radio (NR) RANs. O-RAN Alliance specifications support the split 2 (CU/DU) and split 7.2 (DU/RU) from 3GPP.NR 7.2 split for base stations. The 7.2 split disaggregates base station functionalities into a central unit (CU), a distributed unit (DU), and a radio unit (RU). It connects them together and to a RIC through open interfaces that can consume appropriate data from the RAN and deploy control actions and policies to the RAN.

2. Open vRAN Architecture

2.1. RAN Intelligent Controller (RIC)

Figure 1 shows the Open RAN architecture with RIC platforms and the various open interfaces.

Figure 1. Open RAN architecture with RIC platforms and open interfaces.[4]

The non-real time RIC platform is a part of the service management and orchestration (SMO) platform. The SMO delivers:

- multi-cloud orchestration across regional data centers, national data centers, and public clouds;

- multi-domain orchestration across the RAN and transport and core network domains;

- multi-tenant orchestration across different mobile virtual network operators (MVNOs), enterprises, neutral hosts, and other entities.

With the RIC, the operator can:

- program the network to exercise one or more use cases;

- control the network to achieve a specific objective;

- fine tune the network parameters;

to achieve a certain operator-specified objective function. The objective function may vary depending on the time of the day, location, etc.

The O-RAN RIC architecture divides the necessary functions into SMO, non-RT RIC, near-RT RIC, and applications (xAPPs, rAPPs).

The non-RT RIC is an element of the operator's centralized SMO framework, as defined by the O-RAN Alliance. It hosts the rAPPs that permit the non-RT RIC to exercise one to a few seconds control of RAN elements and their resources. It also provides network data, performance metrics, subscriber data, and AI-based recommendations for network optimization to xAPPs running on the near-RT RIC.

The near-RT RIC platform resides within a telco edge or regional cloud and typically enables network optimization actions that take between 10 milliseconds to one second to complete. It hosts the xAPPs that permit closed loop optimization at fast time scales based on operator-driven intent, and data collected from the network. Similar to the non-RT RIC, the near-RT RIC also utilizes open APIs to connect to xAPPs.

The RIC is architected with northbound and southbound open interfaces for easier integration with partner solutions in the Open RAN ecosystem.

The RIC brings multivendor interoperability, intelligence, agility, and programmability to Open RANs, which enables the onboarding of native and third-party xAPPs and rAPPs to automate and optimize RAN operations at scale. These apps also support innovative use cases that lower mobile operators' TCO and enhance customers' quality of experience (QoE). The rAPPs/xAPPs are available in an Open RAN "app store," to enable operators to create new business models, personalize service experiences, and optimize Capex and Opex.

2.2 Components and Interfaces

Figure 1 shows the four key Open RAN interfaces namely, A1, O1, open fronthaul M-plane and O2, that connect the SMO framework to O-RAN network functions and the O-cloud. Within the logical architecture of O-RAN, the radio side includes near-RT RIC, and CU-based control plane and data plane (O-CU-CP, O-CU-UP), DU, and RU entities.

The non-RT RIC is connected to the near-RT RIC via the A1 interface, while the SMO is connected to both the RAN's CU and DU components using the O1 interface. The rAPPs connect to the non-RT RIC via the R1 interface and over standard APIs. This enables the use of rAPPs that are independent of the RIC implementation (i.e., supplied by third parties).

The near-RT RIC interacts with the non-RT RIC via the A1 interface, and it is connected to the RAN components (CU and DU) via the E2 interface.
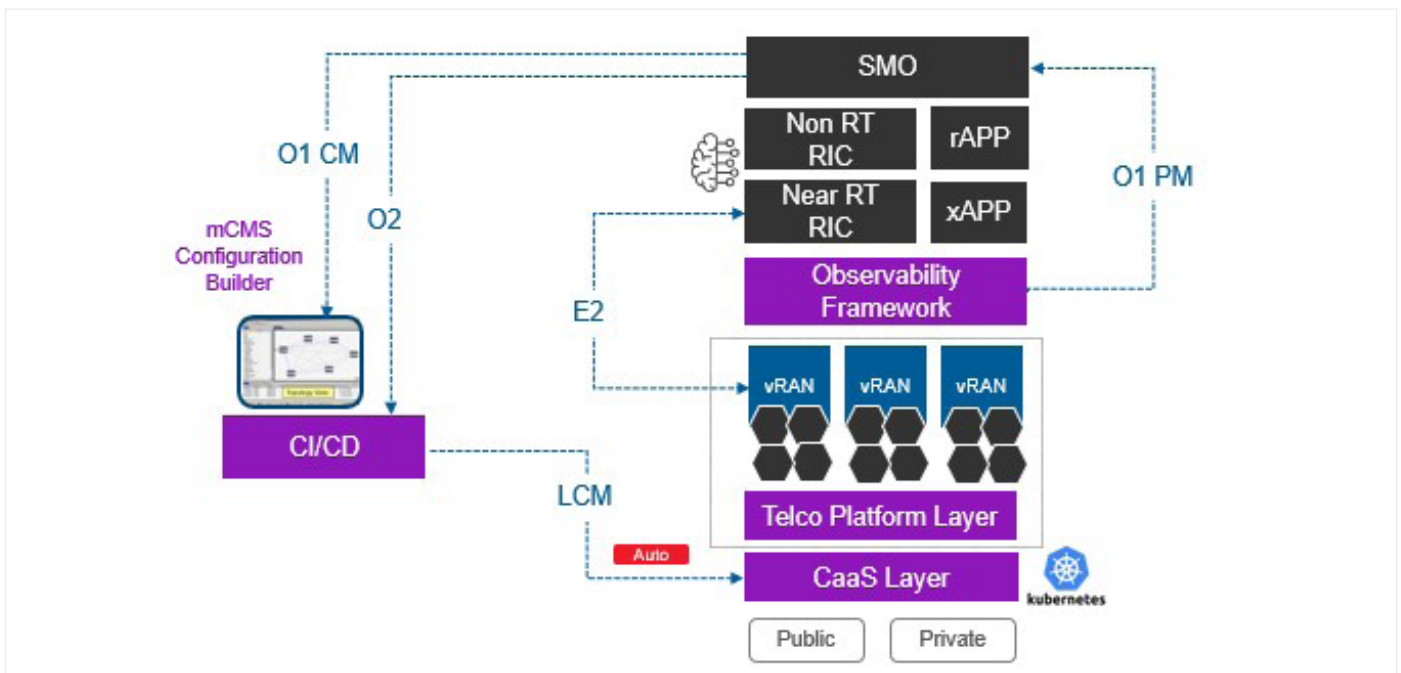

Figure 2. Mavenir O-RIC platform functionality.

3. Mavenir Open RAN RIC (O-RIC)

Mavenir's Open RAN RIC (O-RIC) is comprised of a non-RT O-RIC and a near-RT O-RIC. The O-RIC is a business results tool that enables intent-based decisions by setting granular performance targets—at the cell or even individual UE level—to generate operator business outcomes. These include the generation of additional revenue (for example, by enabling an SLA that generates a premium service), a reduction in network costs (for example through the optimization of resources such as energy savings, spectral efficiency), or the improvement of the end user experience (for example by improving throughput, reducing latency and connection drops, and expanding coverage). The creation of differentiated services is achieved through open APIs that enable intelligent closed-loop network tuning to optimize network performance.

Mavenir's O-RIC offers network intelligence as a service (NIaaS) that provides deep, fine-grained insights about the network and enables building solutions with state of art intelligence. Open RAN networks can now generate new class-of-service revenue streams by adaptively reacting in near-real time to dynamic changes in network, traffic, and load patterns.

The NIaaS signature framework of Mavenir's O-RIC accomplishes this goal by offering predictive intelligence and advanced analytics about the RAN to third-party rAPPs and xAPPs that are integrated with Mavenir's O-RIC. Mavenir's O-RIC allows operators to shift control from the vendor ecosystem to their own ecosystem by delivering an open API framework that allows the implementation of standards-based rAPPs and xAPPs for different business outcomes, such as energy savings, traffic steering, massive MIMO optimization, spectral efficiency improvements, and RAN slice assurance. This is compared with proprietary radio access solutions that limit the carrier's choice to what their vendor has to offer and instead, opens up the network to fine-grained per user data-driven optimization.

With Mavenir's O-RIC, applications can be developed by an operator's in-house team or purchased in an "app store" from any Open RAN-compliant third party creating real business value for next-generation Open RAN networks.

The Mavenir non-RT O-RIC is a containerized application that uses advanced machine learning algorithms to optimize network performance and train ML models using long-term RAN data for dynamic and adaptive policy and control. The non-RT RIC is responsible for setting high-level declarative policies and intents, sending configuration recommendations, and use-case-specific prediction and enrichment information via rAPPs to the near-RT RIC over the A1 interface. The non-RT RIC is hosted in an SMO framework, typically deployed in a centralized cloud, which is responsible for RAN fault, configuration, accounting, performance, and security (FCAPS) operation and orchestration of platform infrastructure resources.

Mavenir further provides O-RAN-compliant near-RT RIC platform, based on Intel hardware, with AI-powered xAPPs, which can control the traffic steering functionality of 5G RAN, a key feature that is responsible for managing the connectivity and mobility of users in the network.

4. Intel technologies for the O-RIC platform

Mavenir's O-RIC software needs a hardware platform that is optimized for both the packet processing needed for the Open RAN software and to run the RIC applications. Servers based on 4th Gen Intel® Xeon® Scalable processors offer up to 56 cores combined with built-in accelerators and other features that are designed to improve Open RAN performance. The CPUs also have new features for managing power and performance, making the best use of CPU resources while also helping to meet sustainability goals.

The network workload-specific 4th Gen Intel® Xeon® Scalable SKUs (N-SKUs) support low-latency, high-throughput applications with optimized system performance. The CPUs support fast DDR5 memory and PCIe 5.0 bandwidth to deliver higher performance[1] for packet processing.

The CPU platform delivers up to twice the capacity at the same power envelope for vRAN workloads versus the prior generation.[6] For vRAN applications, the platform features Intel® vRAN Boost to accelerate forward error correction (FEC) packet processing, freeing up compute cores for other workloads. The integrated Intel vRAN Boost eliminates the need for external FPGAs or ASICs reducing board complexity, cost and the power associated with a discrete board. This platform enables a fully virtualized RAN, enabling maximum flexibility to implement innovations including with AI.

Complementing its processors, Intel continues to provide optimized and scalable FlexRAN™ reference software for vRAN. This software supports L1 functionality in the RAN stack and includes a comprehensive set of features supporting a wide range of 5G deployments, including sub-6GHz, massive MIMO, mmWave, small cell, and private wireless.

The AI/ML models in the Mavenir O-RIC solution predict user quality experience and 5G infrastructure performance KPIs and generate optimized policy/configuration to ensure service level agreements and maximize 5G energy savings. To meet intelligent and automated RIC use-case requirement, it is key to build, tune, and train accurate AI models and deliver performance optimization at the AI pipeline level including data-pre-processing, AI training, and AI inference. Tests conducted by Intel show sub 1 millisecond compute times for the Mavenir AI model inference (see Addendum 1 for workloads and configurations. Results may vary.)

CPU Features Boost AI Development

Intel offers a full-featured development and deployment ecosystem combined with a heterogeneous portfolio of AI-optimized hardware (see Figure 3). Intel's goal is to make it as seamless as possible to accelerate the developer journey from the edge to the cloud. Intel's AI software portfolio includes optimizations for popular frameworks and tools for every stage of the AI/ML process for Mavenir O-RIC solutions.
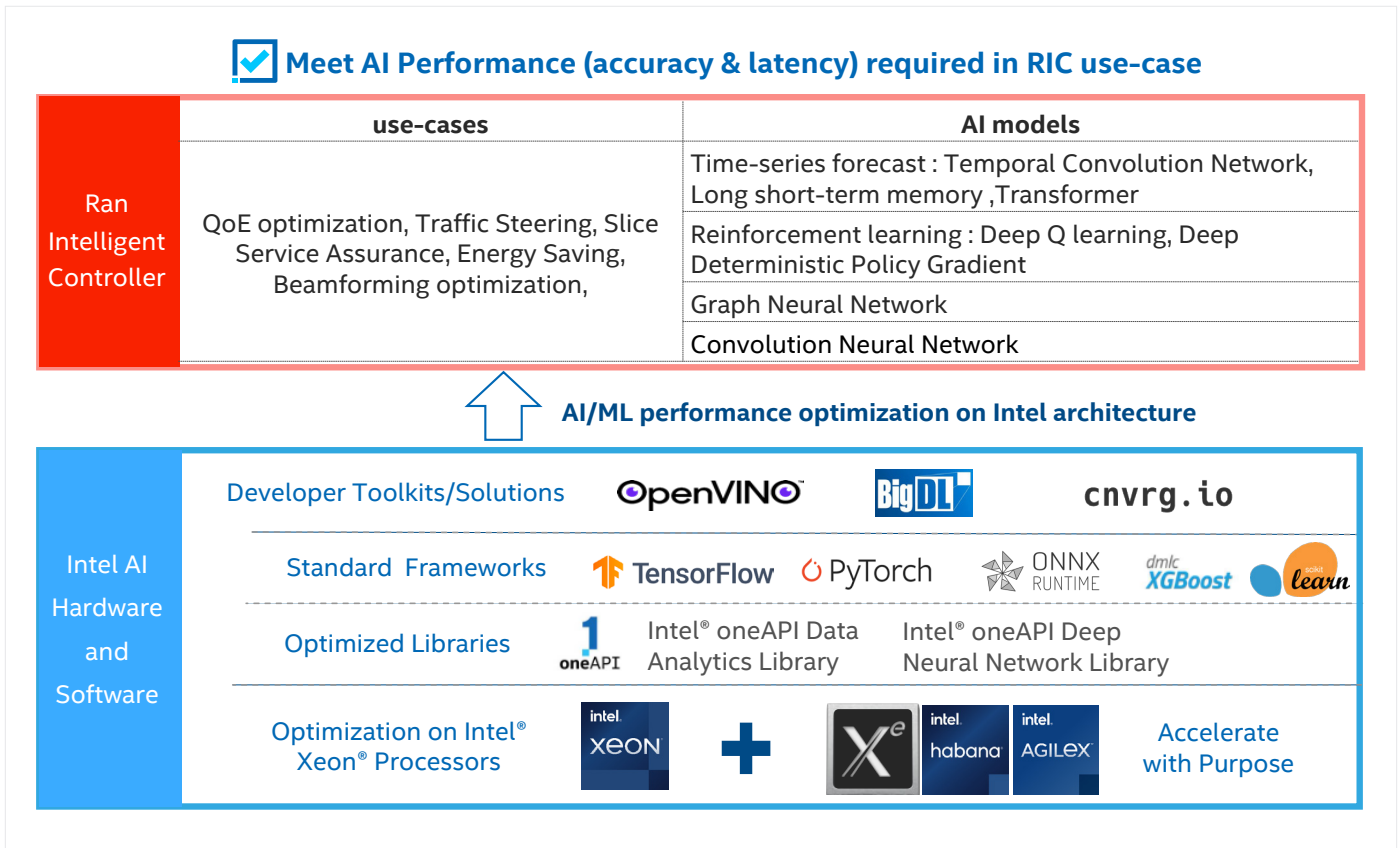
Figure 3. Intel AI hardware and software technologies to optimize AI performance in Mavenir O-RIC solutions.

For AI performance, the 4th Gen Intel Xeon Scalable processors also feature Intel® Advanced Matrix Extensions (Intel® AMX), a built-in accelerator that accelerates the deep learning training and inference that is needed for convolutional neural networks, time series forecasts, and deep reinforcement learning that could be applied in to Mavenir O-RIC use-cases such as RAN energy saving, traffic steering, and RAN optimization.

The Intel® AI Analytics Toolkit (AI Kit) provides tools and frameworks to accelerate the end-to-end analytics and AI pipeline on Intel architecture processors. It provides an Intel-optimized framework for TensorFlow and PyTorch with low-precision tools for high performance AI models training and inference. It achieves acceleration for data preprocessing and machine learning workflows with Python package, Modin, scikit-learn, and XGBoost.

Other AI acceleration features include support for BigDL / Chronos which enables building time-series forecast AI models including data-processing and feature engineering, with the support of more than nine built-in models (TCN, AutoFormer, LSTM, Seq2Seq, NBeats, TCMF, MTNet, ARIMA, Prophet). BigDL / Chronos has a user-friendly API, and AutoML for automatic hyperparameter optimization in a distributed architecture for Mavenir RIC to train and tune highly accurate AI models.

Other hardware features that benefit the Mavenir O-RIC solution include:

Intel® Advanced Vector Extensions (Intel® AVX) for vRAN: Increases vRAN density up to two times compared to the previous generation, with the same power envelope[7]

Intel® Dynamic Load Balancer (Intel® DLB): Provides efficient hardware-based load balancing by dynamically distributing network data across multiple CPU cores as the system load varies.

Intel® Data Streaming Accelerator (Intel® DSA): Drives high performance for storage, networking, and data-intensive workloads by improving streaming data movement and transformation operations.

Intel® QuickAssist Technology (Intel® QAT): Accelerates encryption, decryption, and data compression, offloading these tasks from the processor core to help reduce system resource consumption.

Challenge	Use Case	Business Value
Reducing OPEX and environmental footprint	Energy Savings	Reduce electricity bills and environmental footprint
Automated performance tuning and KPI maintenance	Coverage and Capacity Optimization (CCO)	Fix coverage holes
	Mobility Robustness Optimization (MRO)	Ensure against dropped calls
	mMIMO optimization	Spectrum and capacity savings
Creating new revenue streams	RAN slice SLA assurance	Enable guaranteed SLA
	QoE assurance via traffic split	Provide good user experience
Personalized services and custom experiences	Traffic steering	Targeted KPI improvement (throughput, latency)

Table 1. Challenges addressed by various RIC use cases.

5.0 RIC use cases enabled by AI/ML

The O-RAN Alliance has defined a set of use cases that the xAPPs and rAPPs support. These broadly fall in various categories shown in Table 1, which details the challenges they address.

The use of AI/ML tools is growing in importance and will allow faster RF data processing to aid the non-RT RIC decision-making processes. The non-RT RIC and near-RT RIC allow combining platform telemetry through the O2 interface with application telemetry through the O1 and E2 interfaces. RIC open interfaces provide a platform for innovation and healthy cost competition between xAPP and rAPP developers.

The xAPPs and rAPPs executing in the near-RT RIC and non-RT RIC, respectively, leverage AI/ML techniques to implement the popular use cases shown in Figure 4.

OPEN INNOVATION

RIC provides a platform for open innovation in the RAN space.

So far, ALL RAN features have been developed by only a handful of vendors, stifling innovation for decades.

RIC changes that.

OPERATOR CONTROL

It puts operators in control of their most expensive RAN infrastructure and valuable spectrum assets by giving access to the inner processes in the RAN equipment and define custom algorithms for their own business needs.

FINE-GRAINED DEPLOYMENT

It allows fine-grained deployment of per UE policies to cater to growing business needs of the operators in 5G and beyond.

RAN AUTOMATION

It enables RAN automation to reduce cost and time spent on various age-old processes to manage RAN resources.

PROACTIVE MANAGEMENT

With RIC, operators can proactively manage their network resources to mitigate service degradation and keep quality of experience high for their subscribers.

DIFFERENTIATE SERVICES

It provides a key solution for the operators to differentiate their services among one another, in a highly competitive and crowded RAN services market, a crucial aspect for business success.

Figure 4. Benefit realization from various RIC use cases.

Figure 4 shows the benefit realization from the various, well-established RIC use cases. These include:

- Traffic steering that allows operators to specify different objectives for traffic management such as optimizing the network/UE performance or achieving balanced cell load by distributing UEs between cells.

- Energy savings that permits the use of AI/ML algorithms on the RIC to identify opportunities for cell shutdown during off peak usage times, without degrading user QoE KPIs such as throughput, excessive handover attempts, etc.

- Massive MIMO beamforming optimization that allows operators to optimize network performance and QoS by balancing cell loads or reducing inter-cell interference.

- QoS-based resource optimization that can be used when the network has been configured to provide preferential QoS for certain users, for example when the network has been configured to support end-to-end slices. In this case, the Open RAN network has programmable functionality to ensure resource isolation between slices and functionality to monitor and ensure that slice service level specifications (SLS) are fulfilled.

- Network slice subnet instance (NSSI) which provides resource allocation optimization that allocates resources to the NSSI to support the Open RAN nodes that can be optimized according to service requirements. Multi-vendor slices that address the scenario where vO-DU and vO-CU functions composing each slice is provided from different vendors.

RIC provides automation of actions that would otherwise take too long if done manually. Section 5.1 describes two use cases that demonstrate these automation capabilities to achieve higher network performance or higher user performance. This section provides a more detailed view of traffic steering and energy savings use cases.

5.1 Use case: Traffic steering

Mobility robustness optimization (MRO) aims to dynamically improve the network performance of handovers (HO) to provide improved end-user experience and increased network capacity. The objective is to eliminate radio link failures and reduce unnecessary handovers. MRO is realized by automatically adapting cell parameters to adjust handover boundaries based on feedback of performance indicators. MRO minimizes human intervention in the network management and optimization tasks, thus reducing operator expenditures.

Handover decisions made by the MRO capability, which is built into self-organizing network (SON) functionality, considers only cell level statistics in fine tuning the handover thresholds and hysteresis margins. Current traffic management solutions are usually implemented by relocating users among cells, which highly depends on UE feedback in the measurement report. The statistical characteristics of the radio network and the UE behavior information have not been fully exploited and utilized to enhance the network and user experience performance.

The traffic steering use case demonstrates the ability to flexibly configure the desired optimization policies to proactively manage specific user traffic across different cells and access technologies. A mobility management rAPP provides policy recommendations for handovers and sends them over A1 to the near-RT RIC for policy enforcement.

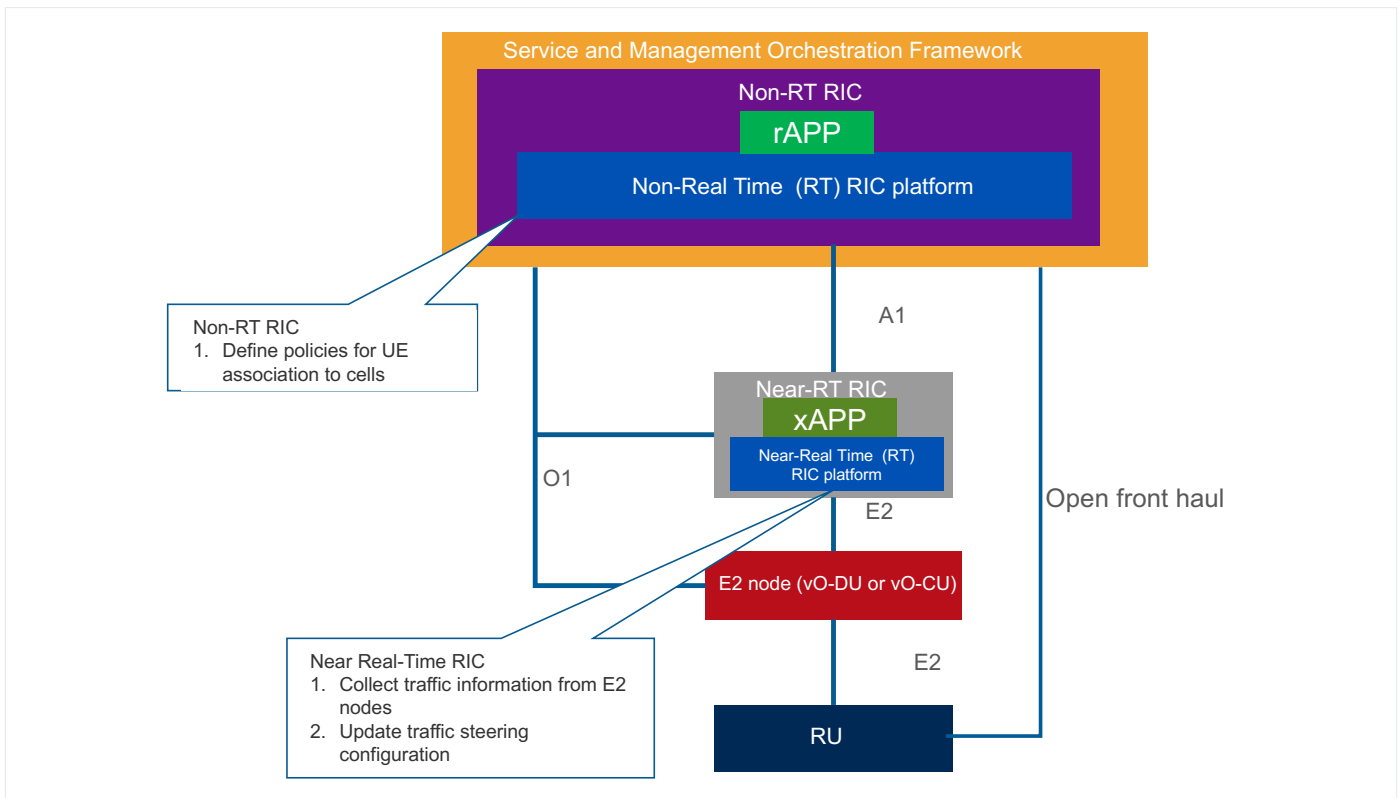Figure 5 shows the schematic of the energy savings use-case.

Figure 5. Traffic steering sub-use case.

A traffic steering xAPP optimizes the cell selection and reselection priority for the UEs to decide the priorities of the absolute radio-frequency channel number (ARFCN) to serve as primary cells. An O-RAN traffic steering xAPP running ML models identifies the optimal target primary cell of individual UEs for handovers. The availability of fine-grained UE information as inputs to the ML/reinforcement learning models enables the traffic steering xAPP to make informed decisions based on learnings from the network to optimize per UE RRM decisions by the E2 nodes (vO-DU, vO-CU).

The traffic steering xAPP hosts a reinforcement learning (RL) agent that receives the UE-specific E2SM-KPM and E2SM-RC indication reports (containing UE context/state information and PMs, serving cells and serving E2 node context and PM information, UE's L3 RRC information for neighbor cells and neighbor cell context information, etc.)

from the E2 nodes (DU, CU). The RL agent in the xAPP then generates a UE-specific handover control action that optimizes the decision of the target cell for the UE that maximize a given KPI target for the UE (such as throughput, latency etc.). Hence, traffic steering enables users to obtain an intent-driven UE specific service.

5.1.1 Benefits of RIC

A simulation evaluation of O-RAN spec-compliant RIC algorithms against SON algorithms and native RRM methods was conducted, with the following configurations.[5]

A dense urban deployment, based on the 3GPP TR 38.913, with M = 1 eNB and N = 7 gNBs. One of the gNBs is co-located with the eNB at the center of the scenario, the others provide coverage in a hexagonal grid. Each node has an independent E2 termination, with reporting periodicity set to 100 ms.
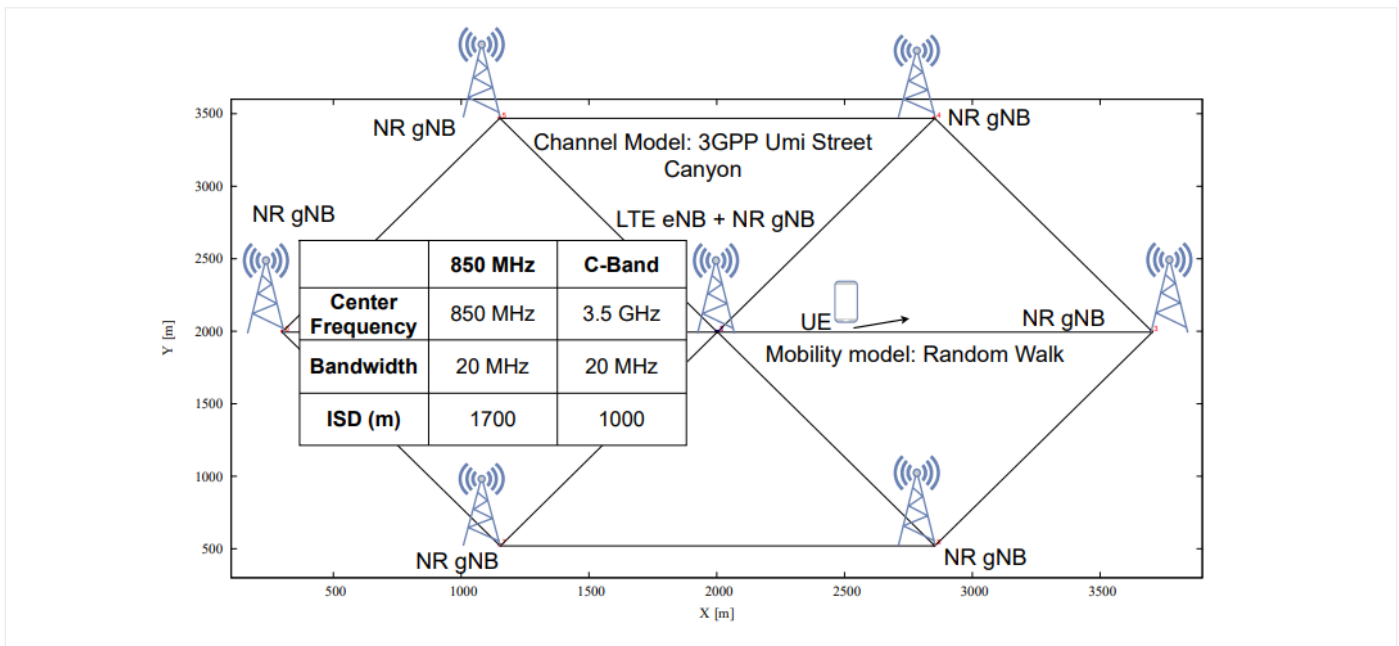
Figure 6. Simulation scenario.[5]

Two different configurations are examined. (i) low band with center frequency of 850 MHz and inter-site distance between the gNBs of 1700 m; and (ii) C-band, with center frequency of 3.5 GHz and inter-site distance of 1000 m. In each configuration, the bandwidth is 10 MHz for the eNB and 20 MHz for the gNBs. The channel is modeled as a 3GPP Urban Microcell (UMi) street canyon channel. The 3GPP NR gNBs use 60 kHz carrier spacing . Dual-connected UEs are randomly dropped in each simulation run with a uniform distribution, and move according to a random walk process with minimum speed Smin = 2.0 m/s and maximum speed Smax = 4.0 m/s. This setup represents the average condition for typical 3GPP scenarios (pedestrian to slow vehicle mobility). We focus on the subset of UEs that are more interested by handovers, rather than, for example, static users, with a random walk model to generalize the mobility through the simulations.

Below are the outcomes obtained.

Throughput

- Average 50% improvement in throughput with RIC compared to SON and RRM.
- Average 30% improvement in 10th percentile throughput with RIC compared to SON and RRM.
- Average 60% improvement in 95th percentile throughput with RIC compared to SON and RRM.

SINR CDF

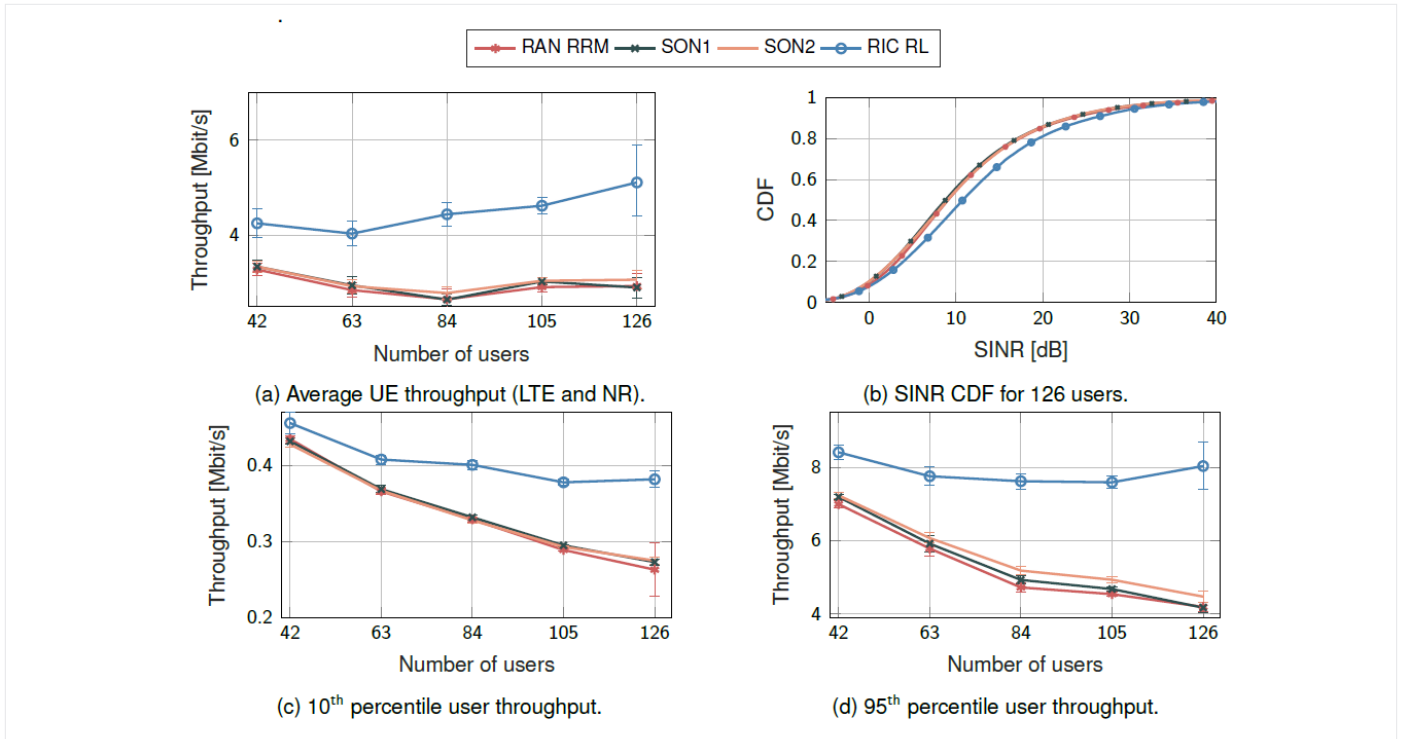- Higher distribution of high SINR UEs with RIC compared to SON and RRM by an average of 4dB.

Figure 7. RIC Traffic steering xAPP simulation – throughput and SINR.

Spectral efficiency

- Average 52% improvement in UE spectral efficiency with RIC compared to SON and RRM.

- Average 35% improvement in cell spectral efficiency with RIC compared to SON and RRM.

PRB utilization and mobility overhead KPI

- Better resource utilization of RIC compared to traditional SON and RRM algorithms.

- Reduced mobility overhead with RIC compared to SON and RRM with the incorporation of cost factor in deep reinforcement learning.
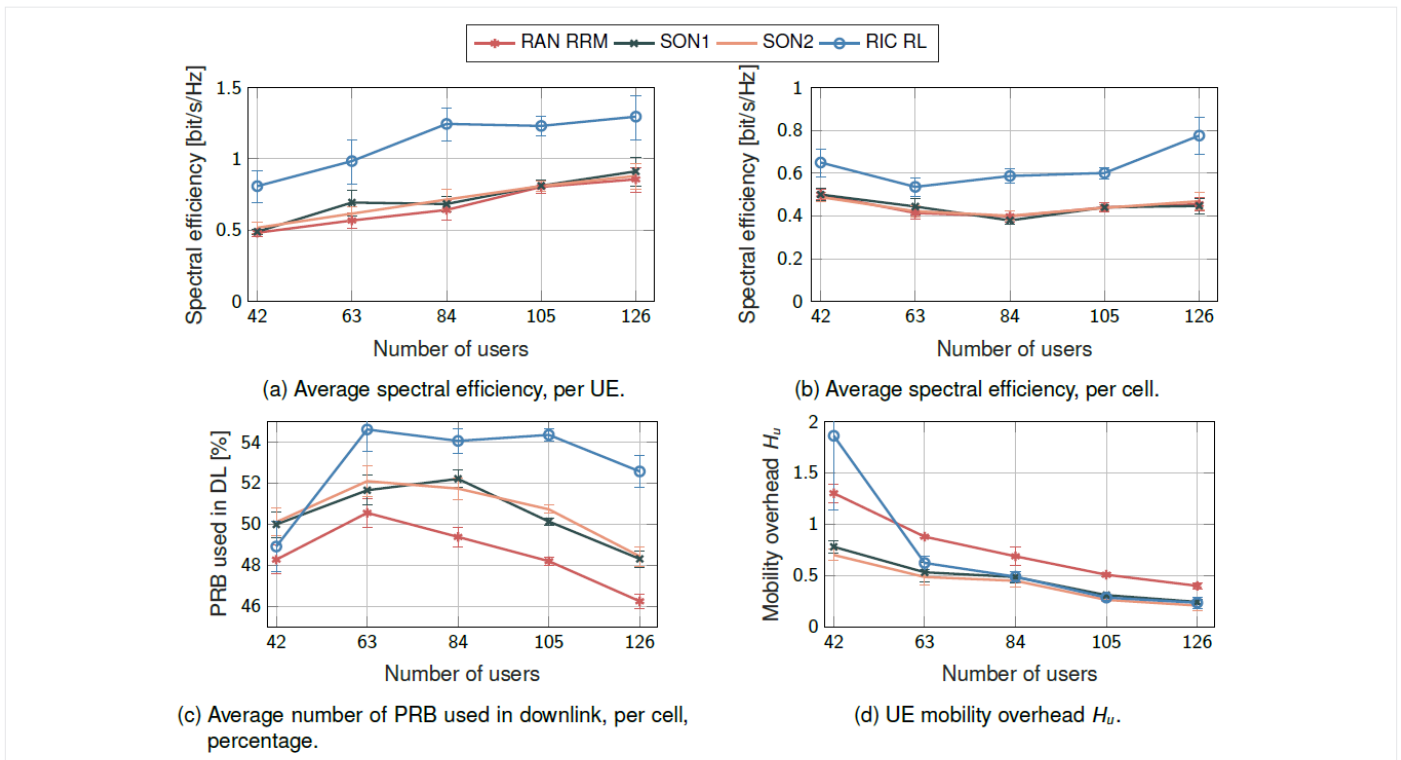
Figure 8. RIC Traffic Steering xAPP – Spectral Efficiency and PRB utilization.

The reinforcement learning models to make these RAN performance gains are accelerated by Intel® Optimization for TensorFlow*, which takes advantage of Intel® AI Engines for the 4th Gen Intel Xeon Scalable processor. The reinforcement learning AI performance results below meet Mavenir AI performance requirements: see Addendum 1 for workloads and configurations. Results may vary.

- AI training latency: 7.72ms with single CPU core
- AI Inference latency: 0.65ms with single CPU core

5.2 Use case: Energy savings with carrier and cell switch off/on

Energy consumption by the RAN is important for 5G network operators as it impacts OPEX. Energy consumption in mobile networks is dominated by the radio access portion of the network[2]. Improvement in energy efficiency is possible using modularization within the system to allow dynamic shutdown of functionalities that are not in use or not required to remain active or on standby for the purpose of synchronization and signaling.

Various methods for energy savings within the radio unit have been identified by Open RAN alliance[3]. One of them, carrier and cell switch off/on, aims at reducing O-CU/DU/RU power consumption by switching off/on one or more carriers or cells. Energy savings gain from carrier and cell switch on/off in an RU is due to reduction in power consumption of the RU hardware components (RF and digital processing units) that are shut down or put into energy savings mode during low traffic load. The O-RAN fronthaul processing, timing and power units continue to function and consume power during the switched off period. The objective of the energy savings use case is stated as follows:

Goal: Maximize the energy efficiency (EE) KPI (given by the ratio of the successfully transmitted data volume from a given cell to the energy consumed by the cell) across all cells in a network.

Control variables: Optimize the subset of cells in a network that can be activated in ES mode and de-activated from ES mode.

Constraint: Minimize the coverage holes in the network. The coverage holes causing RLF/RCEF should not exceed a given threshold.

Due to the conflicting objectives (goal and constraint) and the necessity to model the EE KPI as a function of the subset of cells to be activated in ES mode, AI/ML capabilities are used in the RIC.

This energy savings feature can be invoked in response to changes in traffic load or change in the cell/carrier off/on switching policy. Figure 9 shows a schematic of the energy savings use-case. The non-RT RIC will use the O1 interface to configure the parameters of the O-CU/O-DU and will use the open fronthaul M-plane from the SMO to configure O-RU switch off/on, after triggering handovers to move users to other cells / carriers.

The O1 interface is also used to collect traffic information per cell and per carrier to the SMO as input to the non-RT RIC. rAPPs hosting AI/ML assisted solutions achieve optimized energy savings configurations, such as the switching time that can provide optimal desired tradeoff between energy savings and users' QoS.

Off/on switching is also accompanied with intelligent traffic steering, guided by policies, to ensure service continuity and quality of service for those users connected to the cells that will be shut down during the light load period.
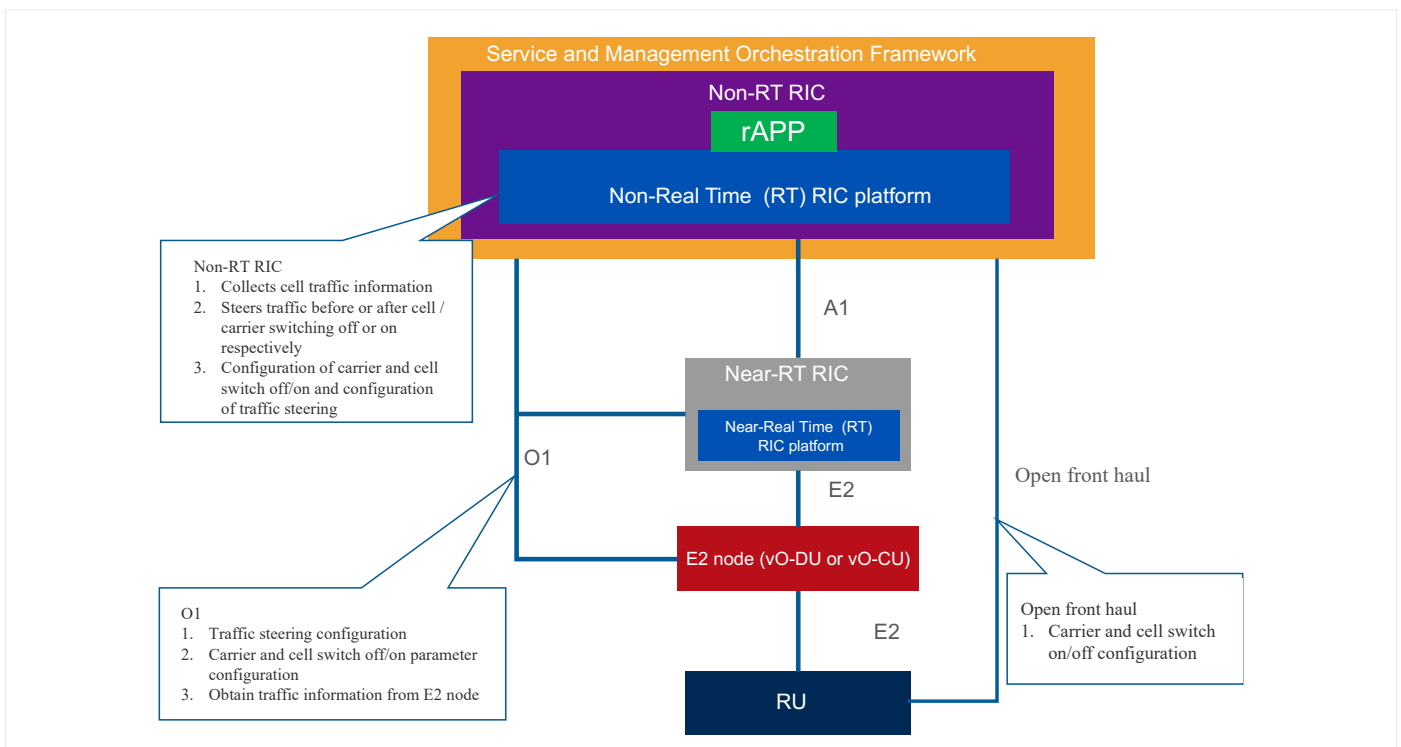

Figure 9. Carrier and cell switch off/on energy savings use case.

5.2.1 Benefits of RIC

An ML target objective is set to maximize energy efficiency KPI namely, data transmitted per unit energy spent. Realistic channel conditions and traffic patterns were generated using UEs, load generators and attenuators. AI/ML assisted solutions based on RL is built into a non-RT RIC rAPP.

Results from the simulation indicate:

- 21% improvement in aggregate EE KPI with energy savings activation when conditions degraded.

- 44% improvement in aggregate EE KPI with energy savings deactivation upon load increase

The proposed solutions will leverage O-RAN open interfaces to support the use case in a multi-vendor disaggregated RAN.

The reinforcement learning algorithm to make these gains is optimized by Intel AI Analytics Toolkit on Intel Xeon Scalable processors to deliver high performance of AI training and inference for RAN energy savings.

6. Conclusion

This white paper outlined the application of the Open RAN with RIC to optimize RAN performance and user experience with various use cases. Access to RAN telemetry data permits the RIC to derive insights to take appropriate closed loop actions in a timely manner, best suited for the use case. These mechanisms are illustrated through the traffic steering use case and the energy savings use case that operate on the near-RT RIC and the non-RT RIC platforms as applications. Key to the effectiveness of RIC use cases is the support from Intel hardware that permits the ability to operate the ML algorithms in a computationally efficient manner.

About Intel

Intel is an industry leader, creating world-changing technology that enables global progress and enriches lives. Inspired by Moore's Law, we continuously work to advance the design and manufacturing of semiconductors to help address our customers' greatest challenges. By embedding intelligence in the cloud, network, edge and every kind of computing device, we unleash the potential of data to transform business and society for the better. To learn more about Intel's innovations, go to newsroom.intel.com and intel.com.

About Mavenir

Mavenir is building the future of networks and pioneering advanced technology, focusing on the vision of a single, software-based automated network that runs on any cloud. As the industry's only end-to-end, cloud-native network software provider, Mavenir is transforming the way the world connects, accelerating software network transformation for 250+ Communications Service Providers in over 120 countries, which serve more than 50% of the world's subscribers.

intel.

References

[1] Open RAN alliance https://www.o-ran.org/specifications

[2] Green Future Networks – Network Energy Efficiency by NGMN Alliance, December 2021
https://www.ngmn.org/wp-content/uploads/211009-GFN-Network-Energy-Efficiency-1.0.pdf

[3] O-RAN.WG1.Use-Cases-Analysis-Report-v09.00.01, March 2023.

[4] O-RAN Architecture Description v08.00, October 2022.

[5] "Programmable and Customized Intelligence for Traffic Steering in 5G Networks Using Open RAN Architectures", A. Lacava, M.Polesi, R.Sivaraj et.al accepted for publication in IEEE Transactions on Mobile Computing, April 2023.

[6] See [n8] at intel.com/processorclaims: 4th Gen Intel® Xeon® Scalable processors. Results may vary.

[7] See [n9] at intel.com/processorclaims: 4th Gen Intel® Xeon® Scalable processors. Results may vary.

Addendum 1

4th Gen Intel® Xeon® Scalable processor based server used in AI training and inference testing:

System	Intel Corporation ArcherCity
Number of nodes	1 node
Baseboard	Intel Corporation ArcherCity
Chassis	Rack Mount Chassis
CPU Model	Intel® Xeon® Platinum 8480+
Microarchitecture	4th Gen Intel Xeon Scalable Processor
Sockets	2
Cores per Socket	56
Hyperthreading	Enabled
CPUs	224
Intel Turbo Boost	Enabled
Base Frequency	2.0GHz
All-core Maximum Frequency	3.0GHz
Maximum Frequency	3.8GHz
NUMA Nodes	2
Prefetchers	L2 HW, L2 Adj., DCU HW, DCU IP
PPINs	08b7d81fa52198c3,08b0d21f7f5ec0df
Accelerators	DLB:2, DSA:2, IAX:2, QAT (on CPU):2, QAT (on chipset):0
Installed Memory	256GB (8x32GB DDR5 4800 MT/s [4800 MT/s])
Hugepagesize	2048 kB
Transparent Huge Pages	madvise
Automatic NUMA Balancing	Enabled
NIC	1x Ethernet Controller I225-LM
Disk	1x 3.6T Samsung SSD 870 QVO 4TB
BIOS	EGSDCRB1.86B.0090.D03.2210040151
Microcode	0xab000310
OS	Ubuntu 22.04.1 LTS
Kernel	5.15.0-43-generic
TDP	350 watts
Power & Perf Policy	Performance
Frequency Governor	performance
Frequency Driver	intel_pstate
Max C-State	9
Tensorflow	2.11
Tested AI Workloads	Mavenir reference learning training and inference for traffic steering xAPP
Tested by Intel as of 04/14/2023	