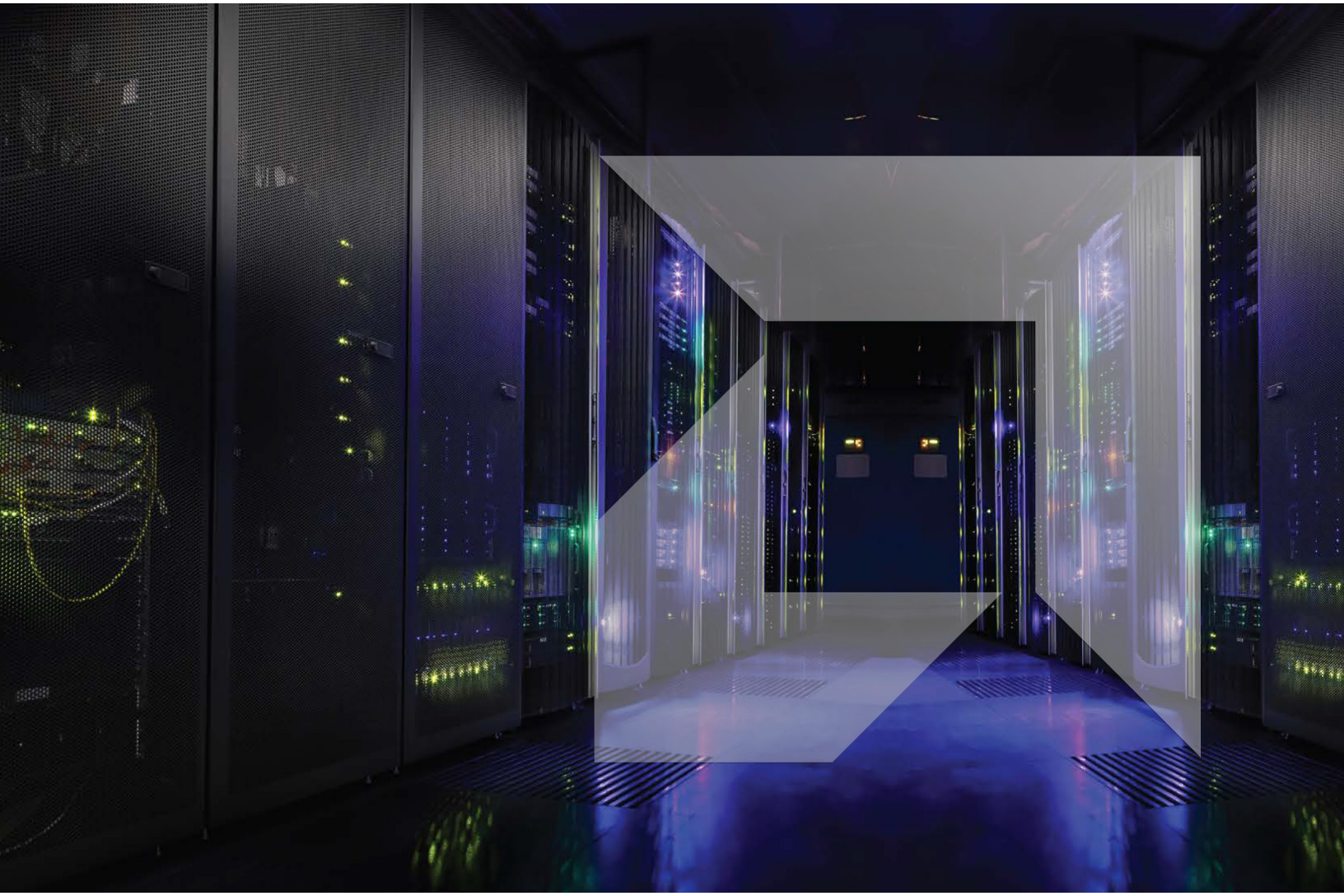
AMD
ALVEO



White Paper | **ADAPTIVE SMARTNICS FOR FUTURE DATA CENTER ARCHITECTURES**

REVISION 1.2023

This white paper is a technical explanation of what the discussed technology has been designed to accomplish. The actual technology or feature(s) in the resultant products may differ or may not meet these aspirations. Each description of the technology must be interpreted as a goal that AMD strived to achieve and not interpreted to mean that any such performance is guaranteed to be fully achieved. Any computer system has risks of security vulnerabilities that cannot be completely prevented or mitigated.

AMD

TABLE OF CONTENTS

EXECUTIVE SUMMARY

In recent years, the system architecture of many network and storage systems has evolved towards a server-centric model. Traditional proprietary purpose-built hardware architectures are seeing a fast decline in usage. These trends started with hyperscale data centers and more recently spread to Telco, Enterprise, and Edge applications. In addition to scalability and resiliency, this homogenized hardware requires agility to morph per deployment and offer workload-specific efficiencies. Workload-specific efficiencies can be achieved if core networking, storage, data movement and security functions are offloaded from the platform host CPU, thus providing more compute resources for the user applications. It is widely established that SmartNICs greatly increase workload efficiencies by offering:

- *Compute intense functions offload from host CPUs*
- *Separation of infrastructure and tenant applications*
- *Adaptability to workload heterogeneity*
- *Inline functionalities*

While today's SmartNICs mostly offer platforms that offload host CPUs, in this paper we showcase many of the emerging use cases of SmartNICs, such as container networking, OVS acceleration, AI/ML, security and others. While existing solutions call for performance with flexibility, these emerging cases increase the value of on-demand accelerators as a critical element of future SmartNIC solutions. Of today's SmartNIC offerings and infrastructure accelerators, FPGA-based implementations such as AMD Alveo™ U25N SmartNICs, Alveo U55C/N accelerator cards, Alveo SN1000 SmartNICs, and Versal™ adaptive SoCs offer the flexibility of dynamic loading of functions. In the future, a wide variety of dynamically loadable functions is expected to be developed by many third-party companies offering these solutions.

INTRODUCTION

SmartNICs have evolved over the past few years in terms of balance between accelerating infrastructure functions vs. tenant applications. Initially, tenant applications were built with a significant amount of infrastructure logic built into them. This made the tenant applications more complex and colossal across all applications. With the advent of more advanced and programmable SmartNICs, the infrastructure logic could be migrated to the SmartNIC, simplifying the tenant application logic as illustrated by the figure below.

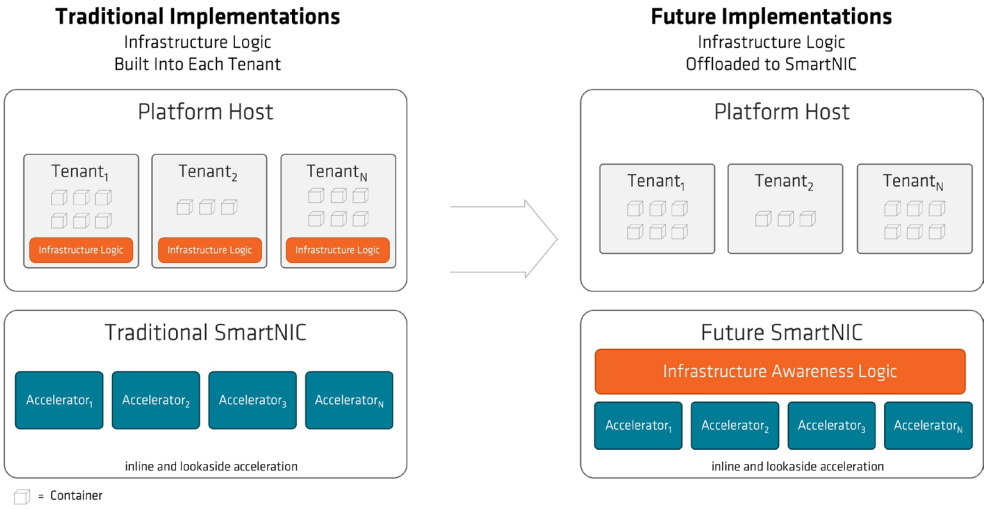


FIGURE 1: INFRASTRUCTURE LOGIC IS INCREASINGLY OFFLOADED TO SMARTNICS

Hyperscalers have demonstrated the OPEX savings of having standard server-based, homogeneous large-scale deployments. As applications grow in scale and are increasingly distributed across tens, hundreds or even thousands of servers, architects are increasingly adding SmartNICs to ensure infrastructure processing doesn't impinge on tenant application processing. Figure 2 lists a summary of the main infrastructure areas that SmarNICs can help offload.

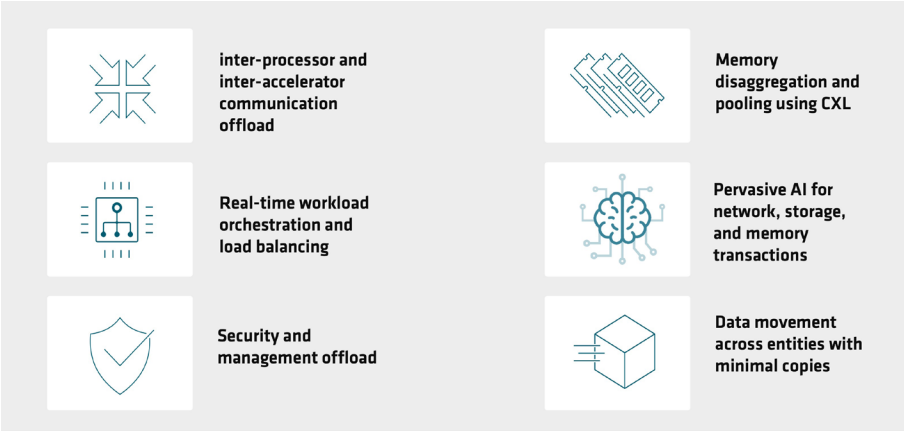


FIGURE 2: CRITICAL FUNCTIONS NEEDED FOR FUTURE DATA CENTERS

In this paper we show six critical functions across cloud, edge, enterprise, and Telco workloads. These emerging SmartNIC applications help homogenous platforms to successfully deliver CAPEX/OPEX savings to end users.

Together with the increased scope of workloads that SmartNICs can offload, workload development and deployment methodologies are changing. Examples include the need to:

a. Facilitate tighter interconnection between computing (both to/from and inter-processor) and other accelerators on the platform
b. Add pervasive AI – monitoring all sensors and transactions by running real-time AI/ML
c. Enable memory disaggregation / pooling architectures using CXL
d. Include hardware based real-time workload orchestration and load balancing
e. Allow data movement with minimal data copies
f. Provide modular root-of-trust and management

INTER-ENTITY COMMUNICATIONS

As workloads span more servers ("scale-out"), inter-processor communication (IPC) is critical to maintaining performance of such workloads. The general trend of scale-out applications introduced chained, containerized lightweight applications that require processing and communications agility. In most data center environments, the amount of east-west traffic (inter-processor) is 3 to 5x greater than that of north-south (line-side) traffic. For example, if the platform supports a NIC bandwidth of 200Gb/s, inter-process communication bandwidth can exceed 1Tb/s.

While today's SmartNICs offload inter-virtual machine communications by offering Open vSwitch (OVS) acceleration, this is not suitable for inter-container communications when the containers are ephemeral (unlike virtual machines). In addition, all the containers within an enclave can share data. In addition, isolation requirements between them is less stringent than requirements between virtual machines. This offers new opportunities for acceleration such as service mesh acceleration in SmartNICs. Containers offload their networking to the proxy or sidecars so applications running in containers are less encumbered with networking tasks. However, today's SmartNICs designs do not yet provide enough performance when running container sidecars and control plane orchestration from Istio, Docker, etc.

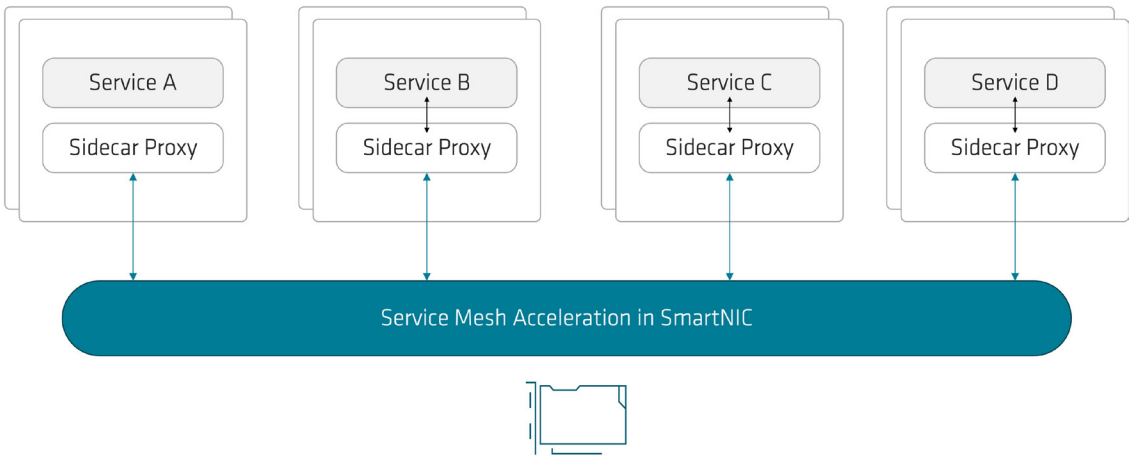


FIGURE 3: SMARTNICS FOR PROXY OFFLOAD AND ACCELERATING CONTAINER NETWORKS

PROXY OFFLOAD

In container networking a proxy-sidecar enhances the functionality of existing containers. It helps move the functions that are unrelated to the application logic in the proxy. This infrastructure separation significantly reduces application complexity. In addition, it increases the agility of applications, enabling porting to different types of infrastructure without specific deployment level assumptions being hard-coded.

Proxy communication, however, tends to exist at higher layers of abstractions such as gRPC or HTTP2. While this may help eliminate going through the network stack, it highlights some performance gaps in today's compute infrastructures in their ability to handle higher layer communications effectively. This requires functions such as HTTP, gRPC, and TLS. For such implementations, there are many functions that need to be accelerated in hardware.

PERVASIVE AI

As infrastructure becomes more heterogeneous and more complex, a programmatic approach to management and security becomes a big challenge. In addition, for cybersecurity use cases like ransomware detection, network anomaly detection, and DDoS prevention, collecting disk access information and performing AI inference on the data helps to detect many of the known attack patterns allowing corrective actions to be taken. Therefore, AI/ML techniques for a multitude of infrastructure use cases across many industry segments is growing exponentially. Tasks ranging from ransomware detection to security anomaly behavior detection to distributed intelligence are all turning to AI/ML for solutions.

AI/ML inspection and inference has been successful when provided with sufficient data sets to initially train the models. SmartNICs can provide AI/ML acceleration for both real-time inline detection and non-real time processing. Ideally suited to AI inference applications, the AMD Versal adaptive SoC portfolio offers built-in AI engine capability along with programmable logic gates.

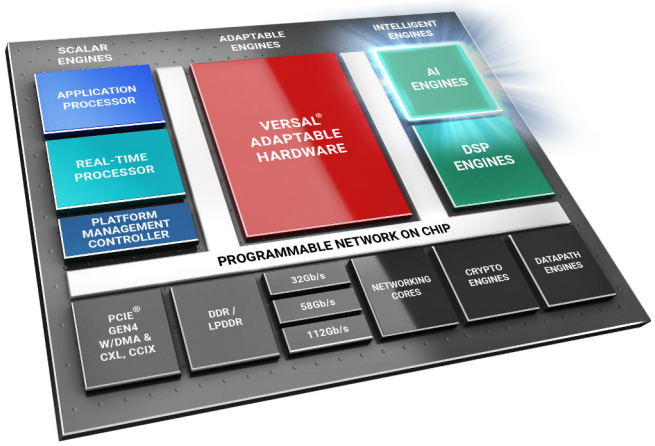


FIGURE 4: VERSAL ADAPTIVE SOCS WITH AI ENGINES FOR NEXT GENERATION SMARTNICS

There are three possible deployment scenarios for SmartNICs to accelerate AI-based infrastructure workloads:

a. A SmartNIC flow classifier helps to select a portion of the network traffic for further behavior anomaly detection. The flow classification and AI engines operate in real-time. Additional statistics collection engines are activated to hand off packet header-based information.
b. A SmartNIC performs as an active data collection agent which selects various metrics to be captured for RX and/or TX packets and then hand off a digest to an AI/ML engine for further inspection.
c. The platform's host CPU runs active data collection agent(s) and, based on higher level intelligent software, selects a portion of that data for the SmartNIC to further inspect.

In summary, SmartNICs will be handling AI/ML workloads natively with integrated, real-time AI inference models.

ACCELERATOR, STORAGE AND MEMORY DISAGGREGATION

With the advent of high-speed and low-latency networking, data center operators are using accelerators, permanent storage, or memory disaggregation to create flexible and scalable architectures. Additionally, Compute Express Link (CXL) and particularly CXL.mem memory solutions are being increasingly deployed. As the CXL 3.0 standard is evolving, FPGAs provide unique flexibility to allow deployed hardware to be upgraded in the field, capable of being updated for the latest standards. AMD UltraScale+™ and Versal™ products offer the advantages of integrated cores and programmable logic for customers who want hardware adaptability without compromising performance.

To enable this disaggregated architecture, dynamic resource allocation, as well as low-latency and simple protocols for end-to-end communication are critical. Figure 5 shows an example of a pooled disaggregated accelerator such as FPGAs and memory.

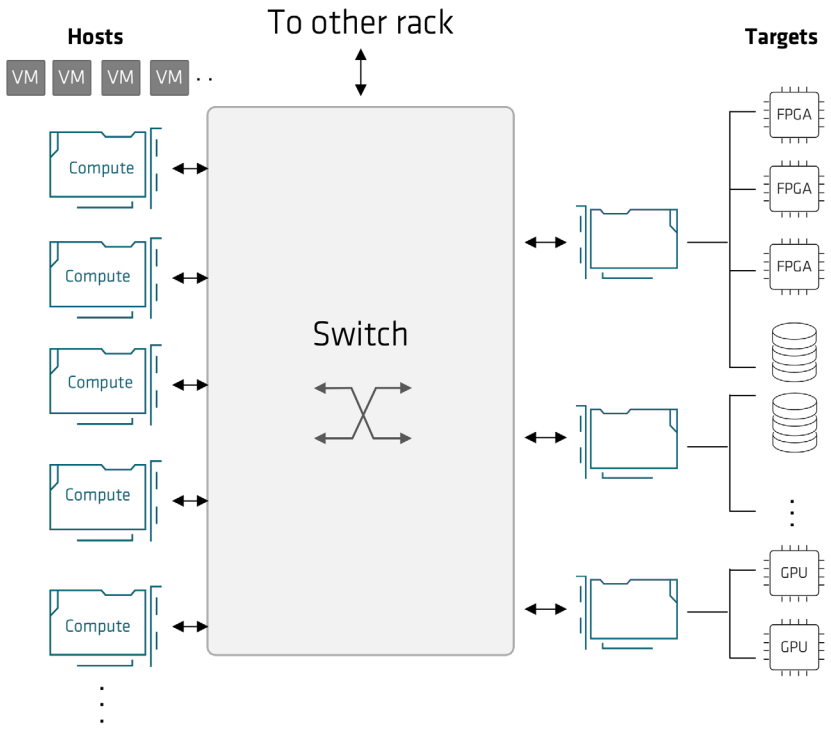


FIGURE 5: ACCELERATOR, STORAGE, AND MEMORY DISAGGREGATION

To enable this disaggregated architecture, there are some essential requirements. First, it must be able to allocate resources dynamically and efficiently. Second, it must support low latency (in the range of a few microseconds). There are several technologies that enable this architecture, such as NVMeoF (NVMe over ROCE, iWarp, Infiniband) and ExpeEther (PCIe over Ethernet).

DATA MOVEMENT OFFLOAD

The number of compute cores in server platforms have been increasing and is soon expected to exceed 100 cores being made available for various workloads. Many threads, processes, virtual machines, and containers are expected to run on these processing cores. For many applications that span across cores or cooperative workloads, the need to hand off data from one core/VM/container to another, data movement between cores is a major challenge. This problem is considerably worse for communication between two tenants that do not trust each other. Data movement requirements for future server platforms will exceed 1Tb/s rates per second. It is therefore inefficient to perform large-scale data movement using CPU cores.

To address these needs, future SmartNICs will be required to offload data copies between tenants. The following figure shows three use cases of data movement:

a. Data movement between containers/VMs and accelerators
b. Data movement between two containers/VMs within a socket or across sockets
c. Data movement between containers/VMs and CXL disaggregated memory

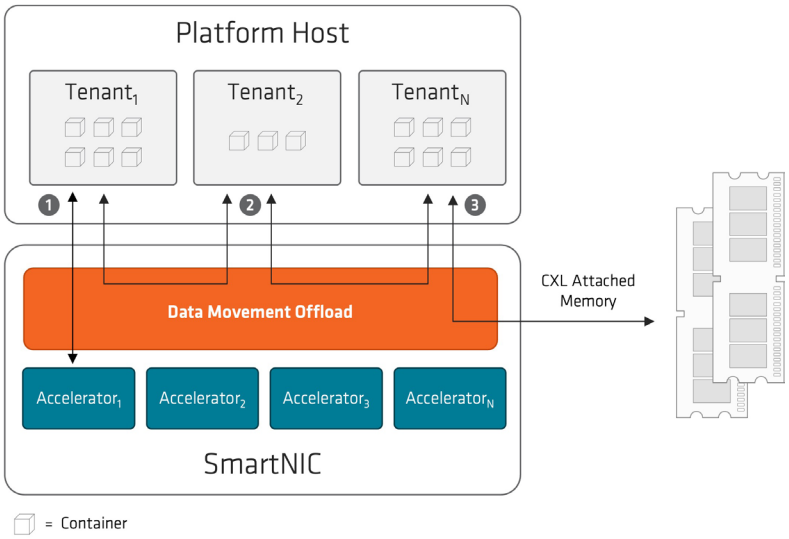


FIGURE 6: SMARTNICS FOR ACCELERATED DATA MOVEMENT

By offloading data movement between entities (containers, VMs, processes), SmartNICs improve energy efficiency, improves core efficiency, enforcement of security and access policies, and offers advanced features of remote atomics used for statistics and other cases.

It is important to note that a data movement accelerator within a SmartNIC needs to have programmable features depending on deployment needs. Another major advantage of routing data movement transactions through a SmartNIC is that all transactions can be monitored in real-time using built-in telemetry engines.

SECURITY AND MANAGEMENT OFFLOAD

Due to the cost of maintenance and upgrade reasons, servers are becoming increasingly modular. Based on this trend, the Open Compute Platform (OCP) has defined modular architectures for the compute, NIC, storage and security functions within a server. This modularity implies that each of the modules needs to be independently manageable and that security of these modules is disaggregated as shown in Figure 7.

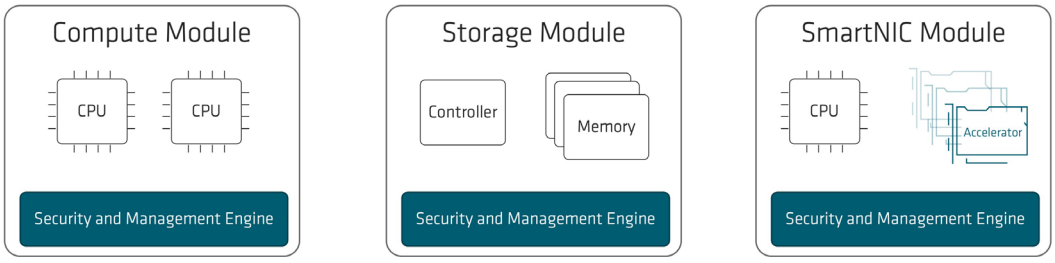


FIGURE 7: MODULAR SOLUTIONS FOR DISAGGREGATED SECURITY AND MANAGEMENT FUNCTIONS

Management functions are traditionally handled by a baseboard management controller (BMC) and security is handled by Root-of-Trust (RoT). SmartNICs already support all aspects of managing this module. Similarly, SmartNICs already support secure boot, key management, and firmware resilience features as part of established security functionality. Given there are many levels of security that can be required, platforms like the portfolio of Alveo SmartNICs from AMD are ideally capable of delivering the desired security level based on specific application needs.

AMD ADAPTIVE SMARTNICS AND NETWORK ACCELERATORS

The following figure illustrates the underlying architecture of AMD adaptive SmartNICs.

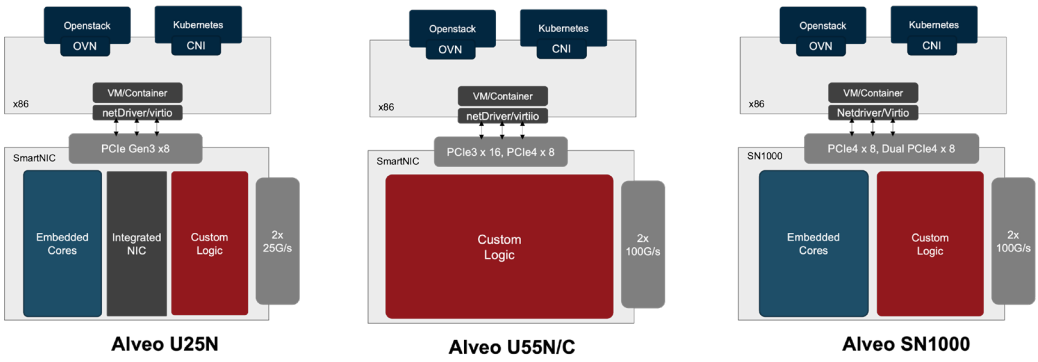


FIGURE 8: ALVEO ADAPTIVE SMARTNIC AND NETWORK ACCELERATOR ARCHITECTURES



Alveo Card	U25N	U55N, U55C	SN1000
Ports	2x 25Gb/s	2x 100Gb/s	1x 100Gb/s or 2x 100Gbps
Control Plane	On Host (internal proxy host on Arm® processor)	On Host	On Arm or Host
Data Plane (Acceleration)	Programmable Logic	Programmable Logic	Programmable Logic
Basic NIC	ASIC/Hard Logic	Programmable Logic	Programmable Logic
Lookup memory	Cache + DDR for lookup	Cache + HBM (HBM for high-speed look-up)	Cache + DDR for lookup
Embedded CPU	Quad-Core Arm Cortex®-A53	-	16-Core Arm-Cortex A72
OpenNIC/Shell	Open network shell	OpenNIC	OpenNIC

TABLE 1: AMD ALVEO SMARTNICS AND ACCELERATOR CARDS

Here the Alveo U55N/C network accelerator cards and Alveo SN1000 SmartNIC are provided with an OpenNIC shell, allowing developers to quickly make use of the underlying features. The OpenNIC project provides an FPGA-based NIC platform for the open-source community to contribute industry expertise. It consists of two components: a NIC shell and software drivers. The NIC shell is a pre-developed design (written in RTL) that configures the AMD FPGA with essential SmartNIC functionality. It currently supports several of the network-enabled Alveo accelerator cards. See https://github.com/Xilinx/open-nic for more information. The software includes both Linux and DPDK drivers. Figure 10 shows the architecture contained in the NIC shell.

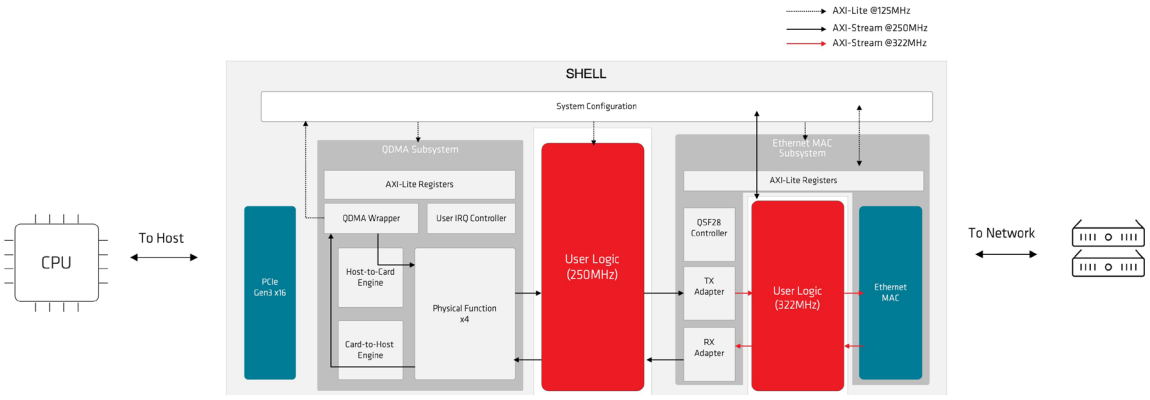


FIGURE 10: OPENNIC SHELL DESIGN

Figure 11 illustrates a hardware programmable and energy efficient SmartNIC, using the Alveo U25N platform as an example.

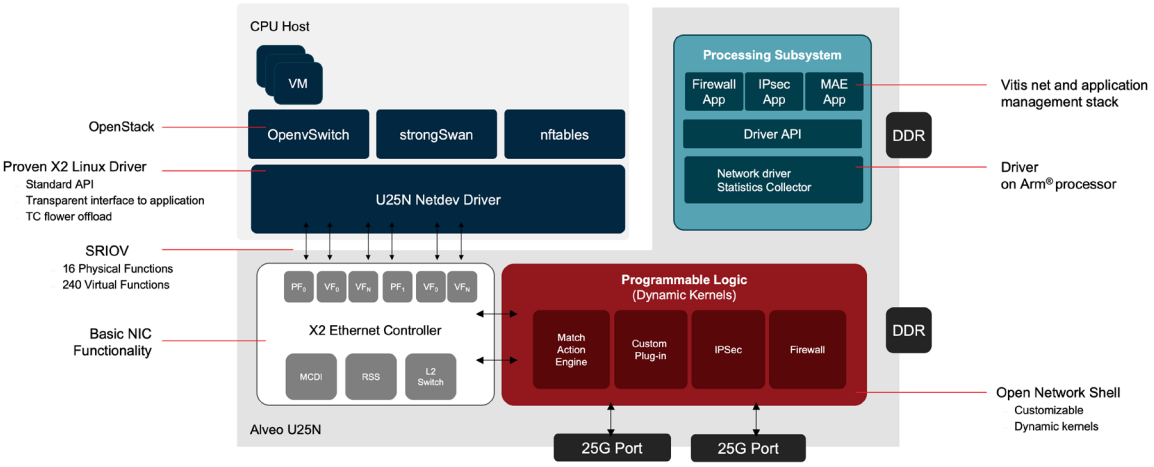


FIGURE 11: ALVEO U25N SMARTNIC BLOCK DIAGRAM

Figure 12 illustrates performance advantages of OVS offload for average packet sizes for a vRouter. As shown in the results, OVS offloaded in U25N SmartNIC shows more than 4 times higher performance than the CPU version of OVS, and uses only 30% of the CPU[i].

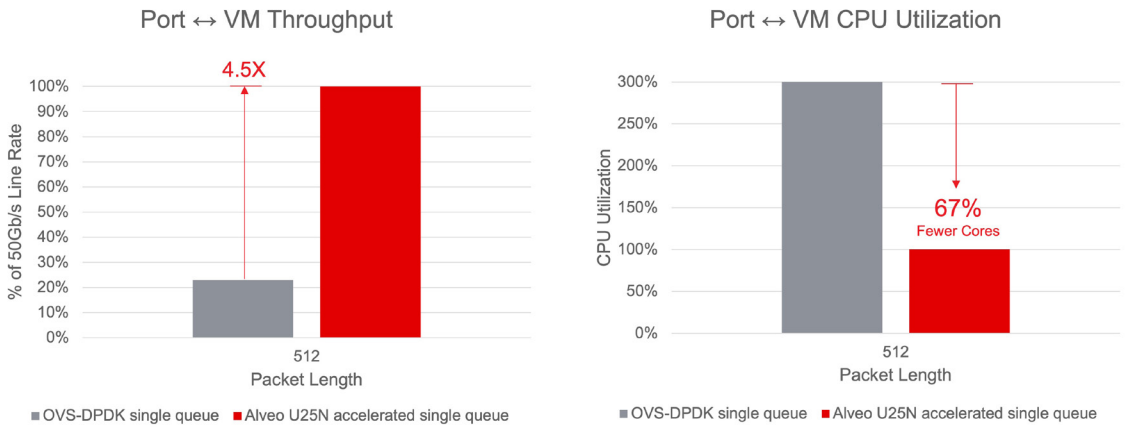


FIGURE 12: ALVEO U25N SMARTNIC ACCELERATED OVS-DPDK VS. SOFTWARE IMPLEMENTATION[i]

SUMMARY

SmartNICs are continuing to evolve and will lead to advancements in data center capability. The use cases presented in this paper highlight that a fixed solution will not fit all deployment needs. The ideal solution involves deployment-specific accelerators leading to cost and power efficiencies. An adaptive solution, such as those provided by AMD have flexibility to enable new data center architectures and be deployed efficiently at scale. We invite industry partners to collaborate with AMD to build the next generation of dynamically loadable accelerators and deploy proof-of-concepts for the various use cases presented in this paper.

REVISION 1.2023

www.xilinx.com/network-acceleration

[i] An average of 4.5x the performance [averaging 3.5x or 353% faster performance], using 67% fewer CPU cores using Alveo U25N SmartNIC with acceleration using driver v8.3, Open vSwitch v3.0.0 and DPDK v22.07 (512 Packet Length) than with same version of Open vSwitch and DPDK without acceleration. ALV-001 Testing conducted by AMD AECG DCCG Data Center Solution Architect Team as of 10/26/2022 on the Alveo U25N SmartNIC, on a test system comprising Model: Dell R740, Dual Sockets server, CPU: Intel® Xeon® Gold 6134 CPU @ 3.20GHz 8cores/16 HW threads, RAM: 96GB@2400 MT/s, HDD: 2.0TB, OS version: 5.15.0-46-generic #49~20.04.1-Ubuntu. PC manufacturers may vary configurations, yielding different results. Performance may vary based on use of latest drivers.