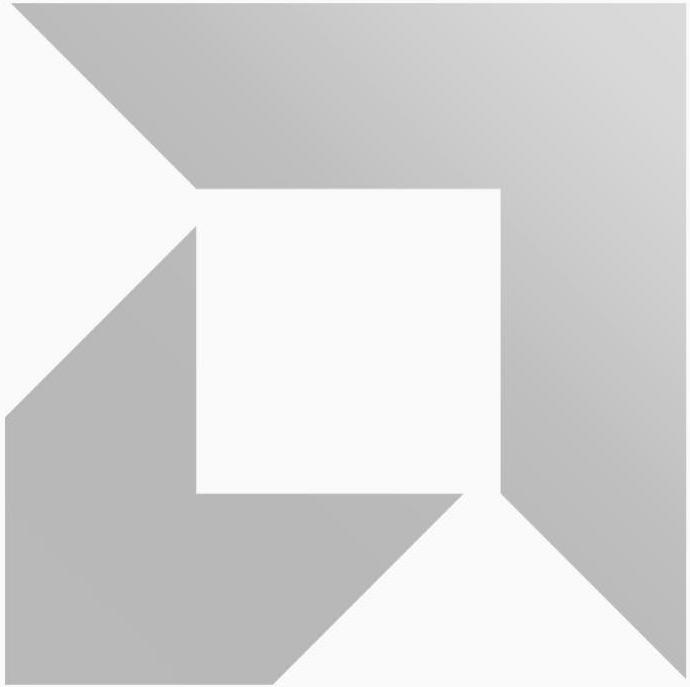


SmartNICs – P4's ~~Final~~ Latest Frontier

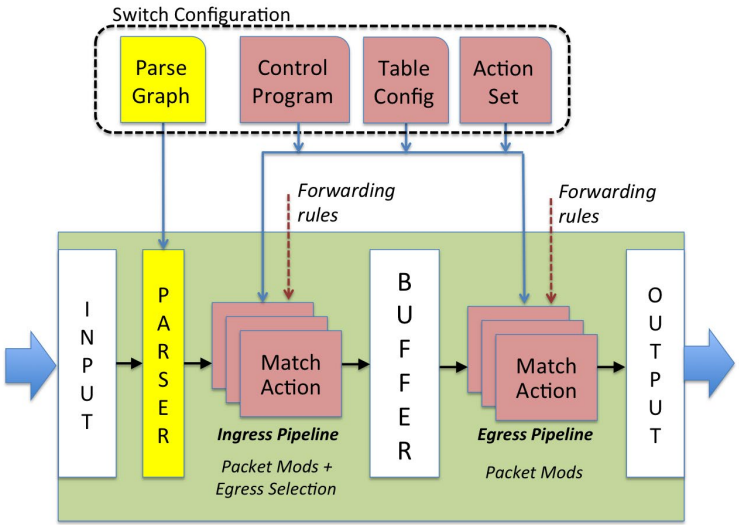
Challenges and Opportunities Ahead

Mario Baldi
Fellow, AMD Research and Advanced Development

7th European P4 Workshop (EuroP4 '24)

AMD
together we advance_

P4: The Origin



P4: Programming Protocol-Independent Packet Processors

Pat Bosshart†, Dan Daly*, Glen Gibb†, Martin Izzard†, Nick McKeown‡, Jennifer Rexford**, Cole Schlesinger**, Dan Talayco†, Amin Vahdat¶, George Varghese§, David Walker**
†Barefoot Networks *Intel ‡Stanford University **Princeton University ¶Google §Microsoft Research

ABSTRACT

P4 is a high-level language for programming protocol-independent packet processors. P4 works in conjunction with SDN control protocols like OpenFlow. In its current form, OpenFlow explicitly specifies protocol headers on which it operates. This set has grown from 12 to 41 fields in a few years, increasing the complexity of the specification while still not providing the flexibility to add new headers. In this paper we propose P4 as a strawman proposal for how OpenFlow should evolve in the future. We have three goals: (1) Reconfigurability in the field: Programmers should be able to change the way switches process packets once they are deployed. (2) Protocol independence: Switches should not be tied to any specific network protocols. (3) Target independence: Programmers should be able to describe packet-processing functionality independently of the specifics of the underlying hardware. As an example, we describe how to use P4 to configure a switch to add a new hierarchical label.

1. INTRODUCTION

Software-Defined Networking (SDN) gives operators programmatic control over their networks. In SDN, the control plane is physically separate from the forwarding plane, and one control plane controls multiple forwarding devices. While forwarding devices could be programmed in many ways, having a common, open, vendor-agnostic interface (like OpenFlow) enables a control plane to control forwarding devices from different hardware and software vendors.

multiple stages of rule tables, to allow switches to expose more of their capabilities to the controller.

The proliferation of new header fields shows no signs of stopping. For example, data-center network operators increasingly want to apply new forms of packet encapsulation (e.g., NVGRE, VXLAN, and STT), for which they resort to deploying software switches that are easier to extend with new functionality. Rather than repeatedly extending the OpenFlow specification, we argue that future switches should support flexible mechanisms for parsing packets and matching header fields, allowing controller applications to leverage these capabilities through a common, open interface (i.e., a new "OpenFlow 2.0" API). Such a general, extensible approach would be simpler, more elegant, and more future-proof than today's OpenFlow 1.x standard.

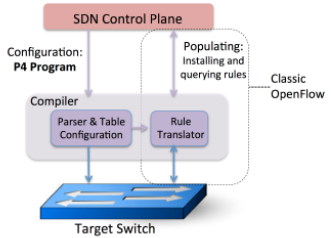
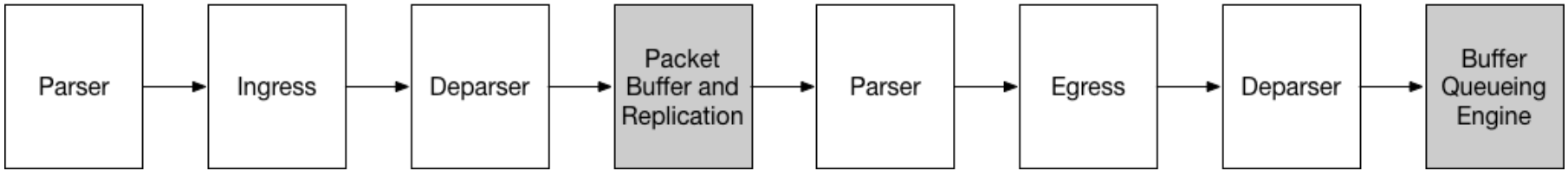


Figure 1: P4 is a language to configure switches.

SIGCOMM 2014

P4 Community – Portable Switch Architecture (PSA)



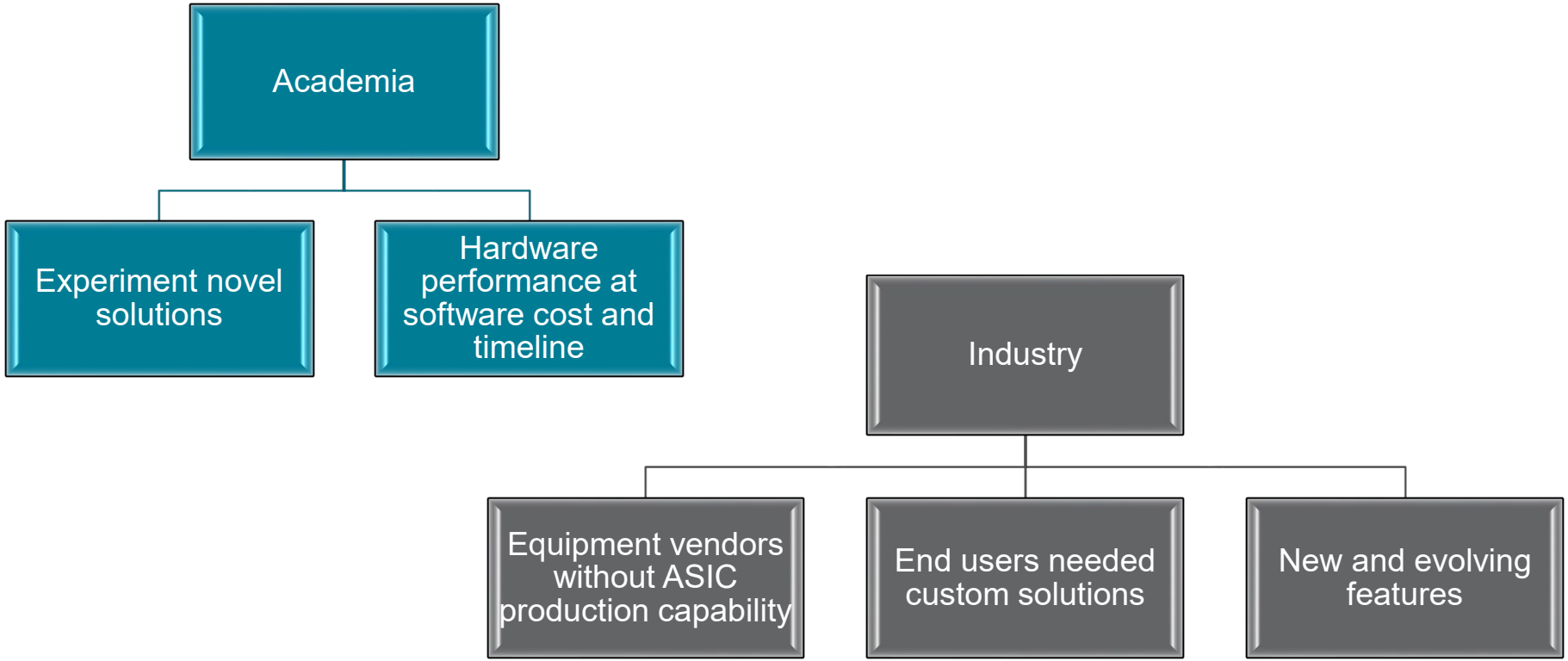
p4.org Architecture Working Group Specification: **P4$_{16}$ Portable Switch Architecture (PSA)**
https://p4.org/p4-spec/docs/PSA-v1.2.html#sec-target-architecture-model

The First Available Products: Indeed Switches

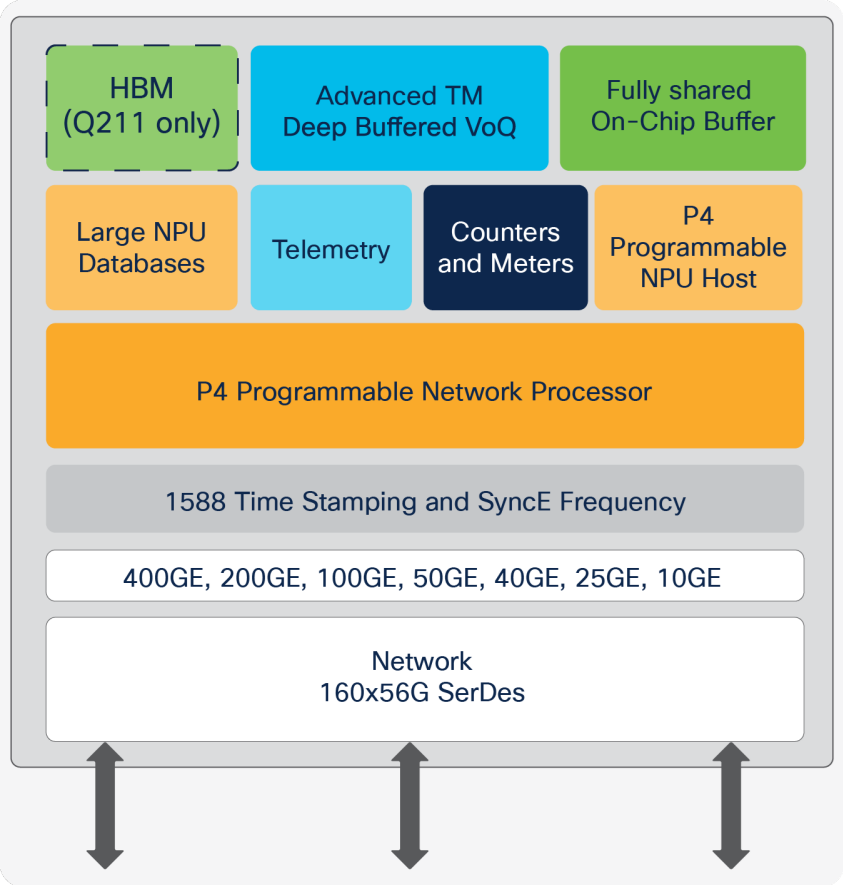
Intel/Barefoot Tofino

https://www.intel.cn/content/www/cn/zh/products/network-io/programmable-ethernet-switch.html

Lots of Interest



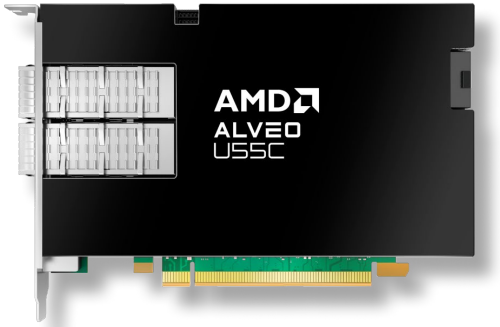
More Switching Products Followed

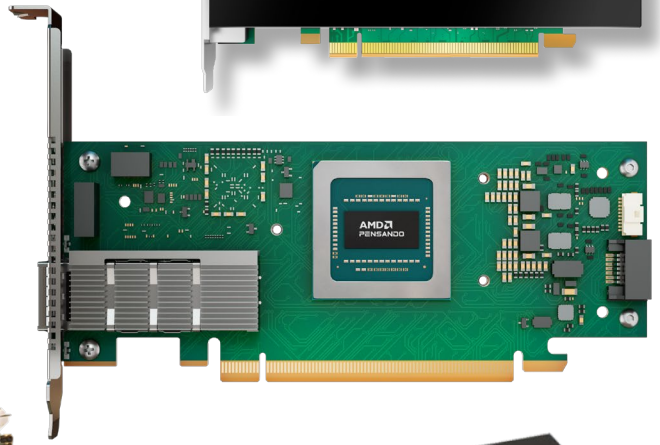


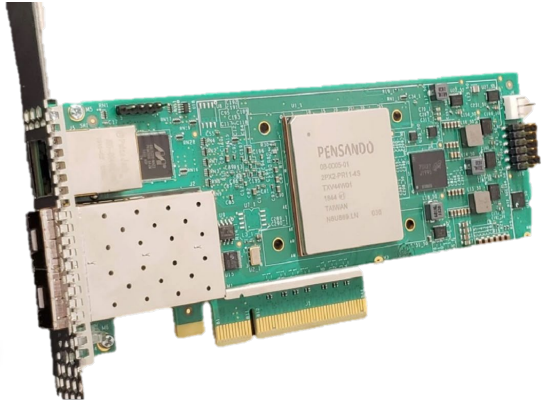
Cisco Silicon One

https://www.cisco.com/c/en/us/solutions/collateral/silicon-one/datasheet-c78-744834.html

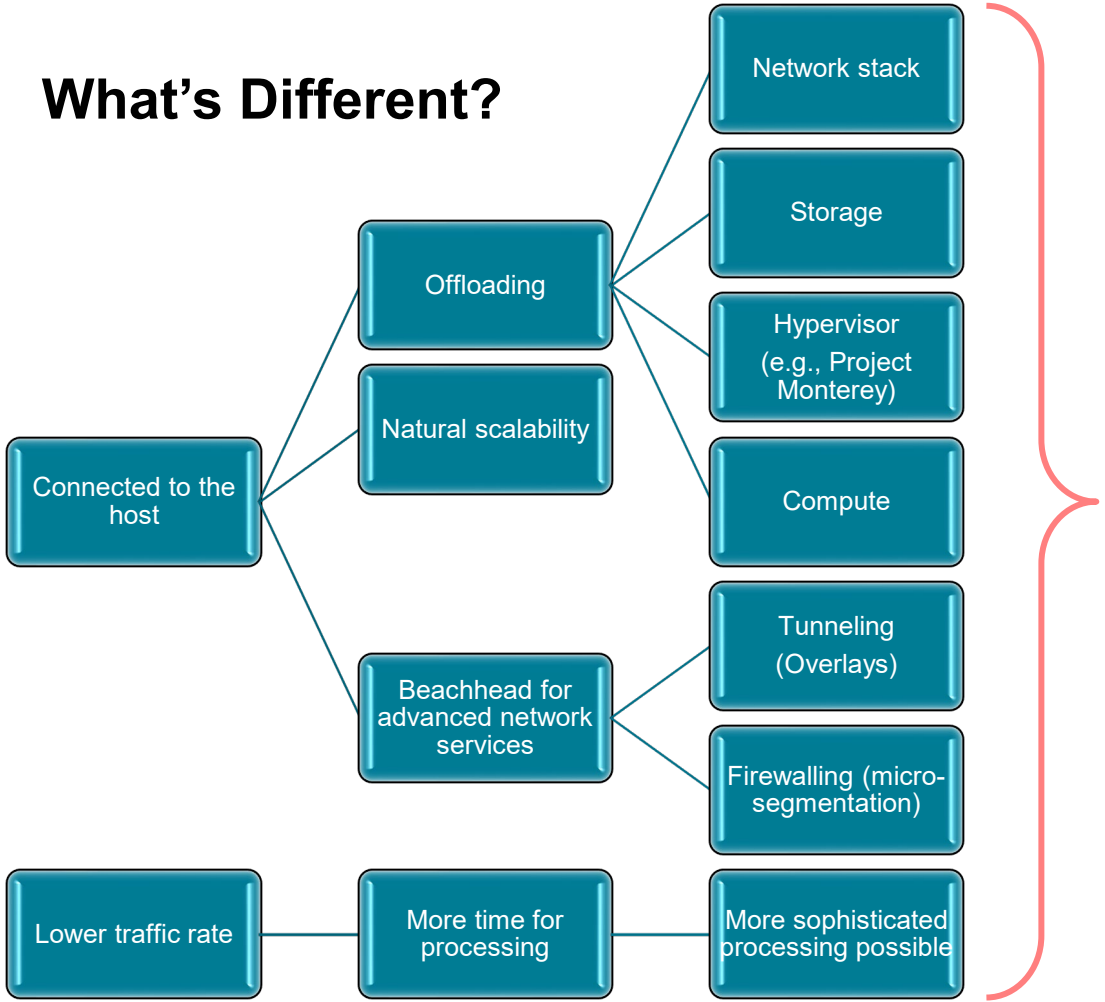
Programmable NICs Make Their Appearance

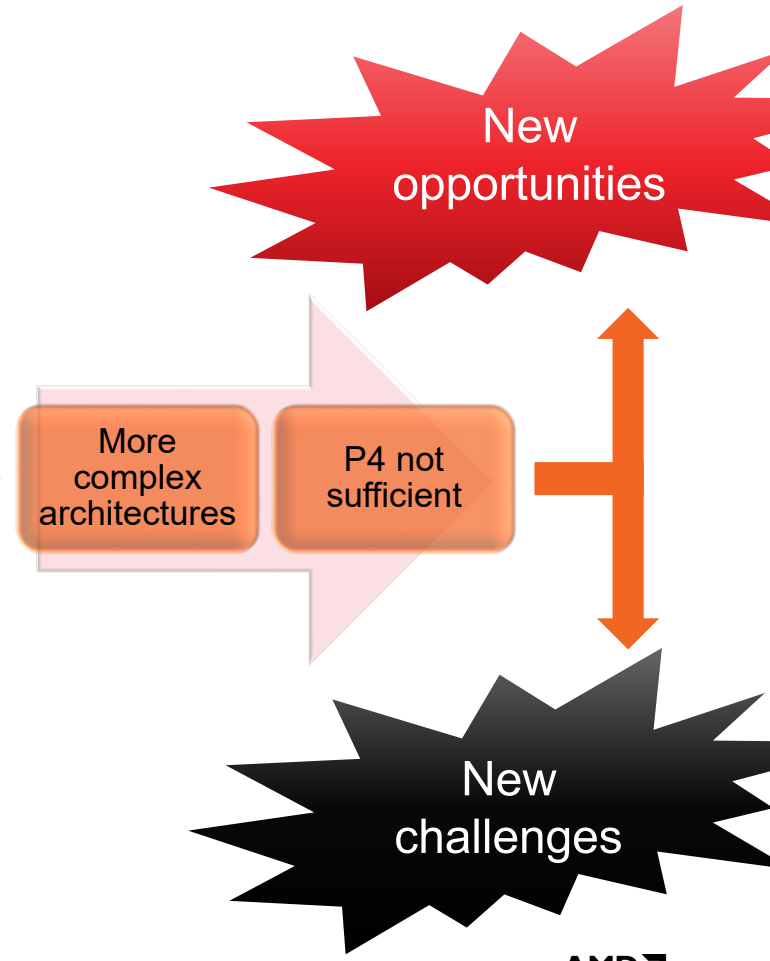




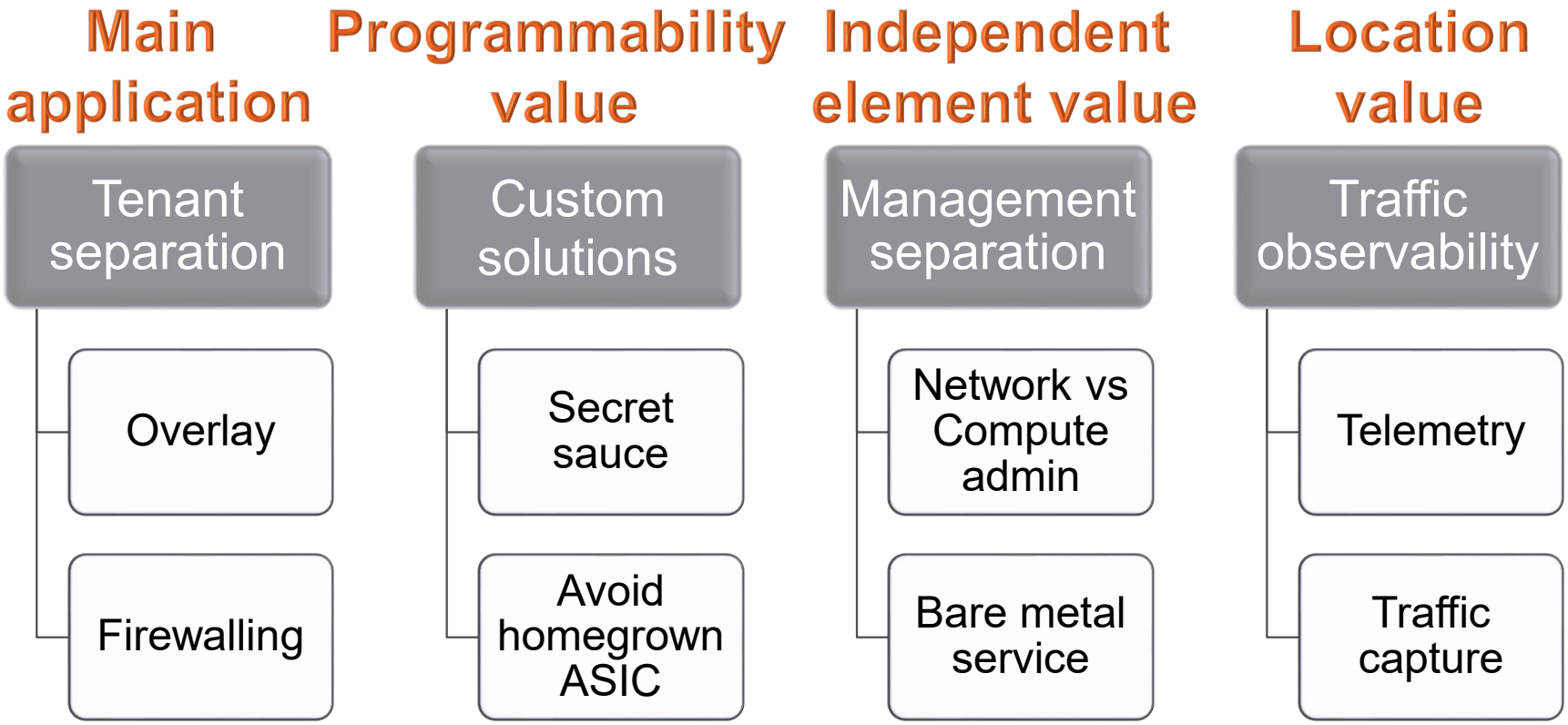


What's Different?

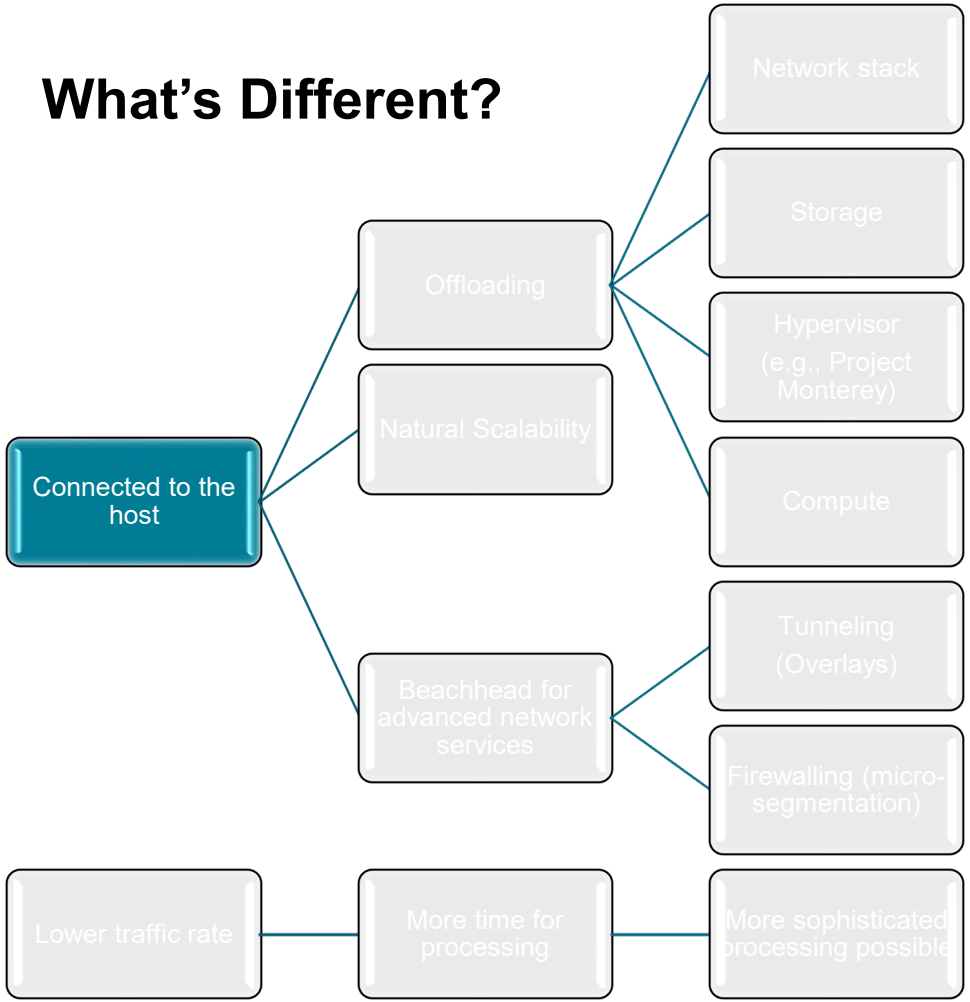




Big Fans: Cloud Providers and Hyperscalers



What's Different?



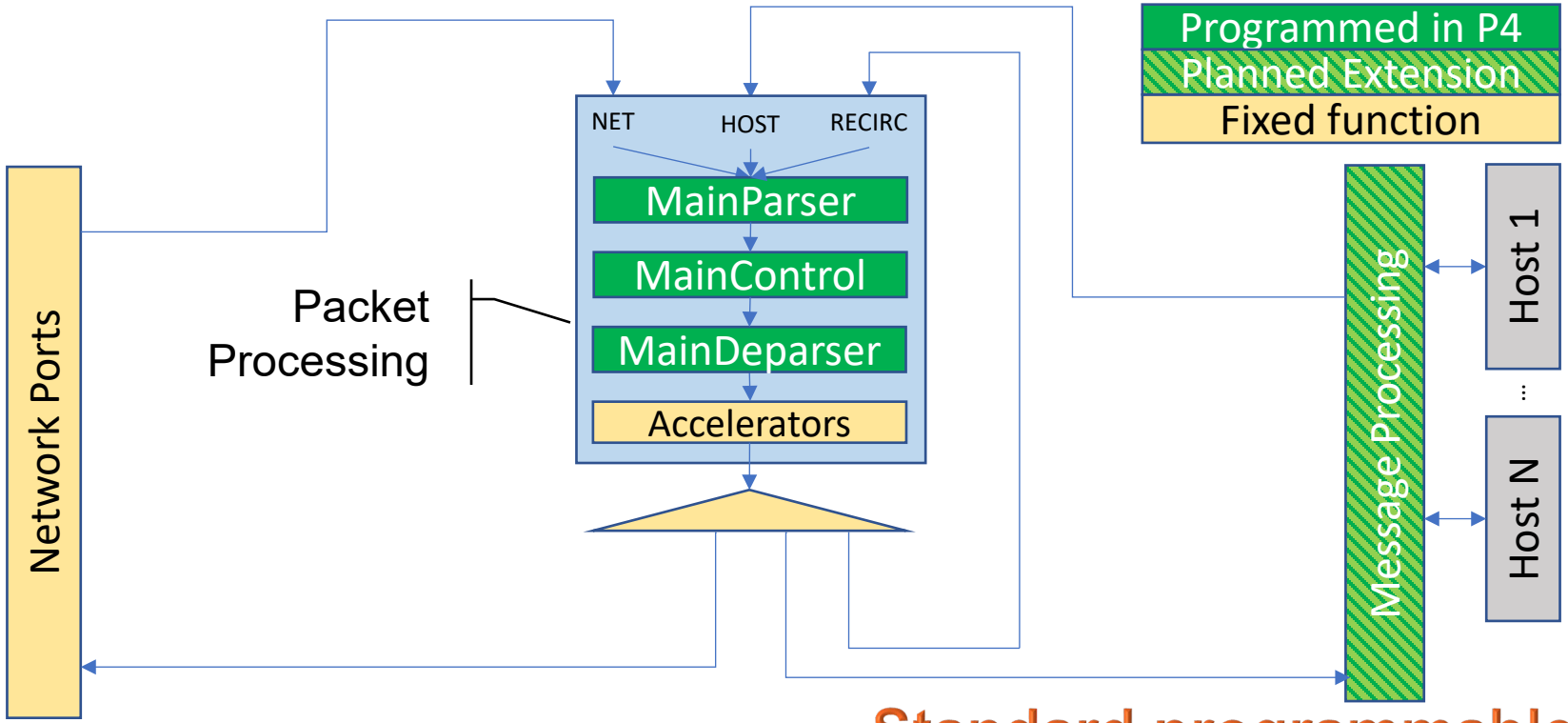
New opportunities

More complex architectures

P4 not enough

New challenges

Portable NIC Architecture (PNA) v0.7



p4.org Architecture Workgroup Specification
https://p4.org/p4-spec/docs/PNA-v0.7.0.html#sec-introduction

Standard programmable interface to the host

PNA Novelties

Add-on-miss

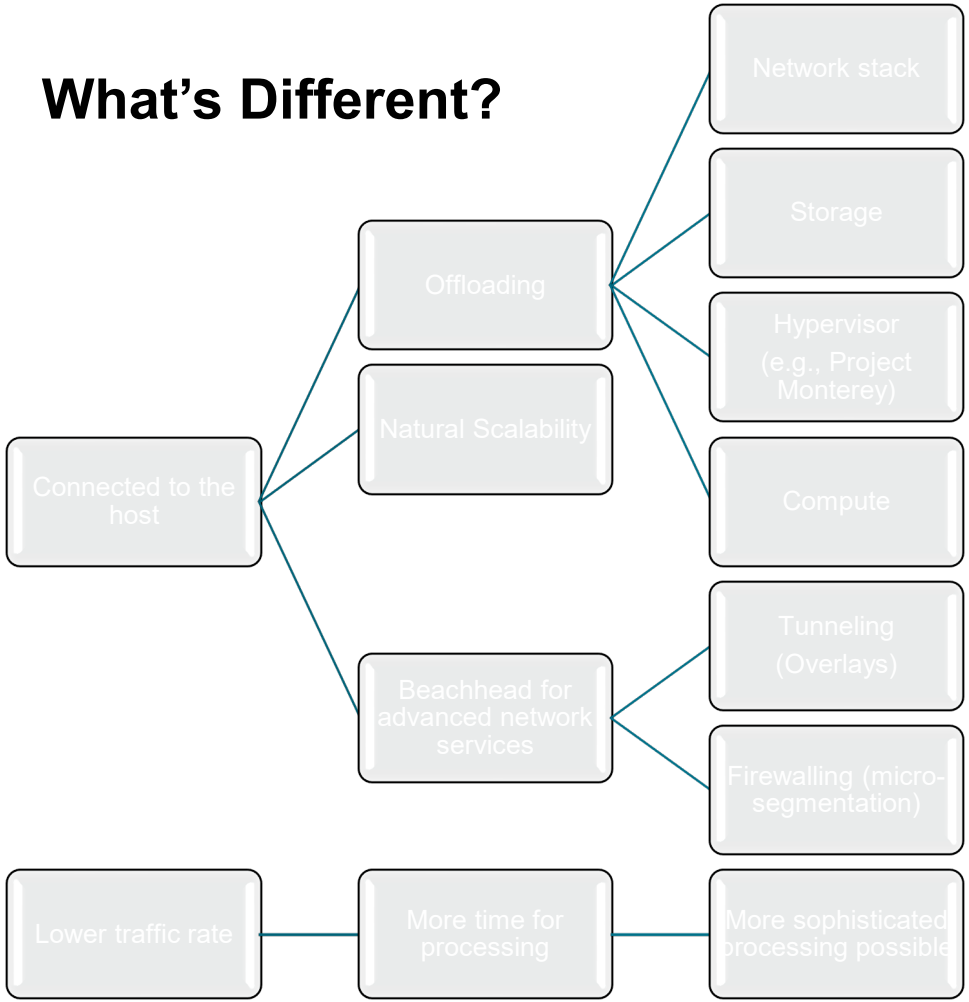
- Create a new entry in a table directly in the data plane
- Without involving the control plane

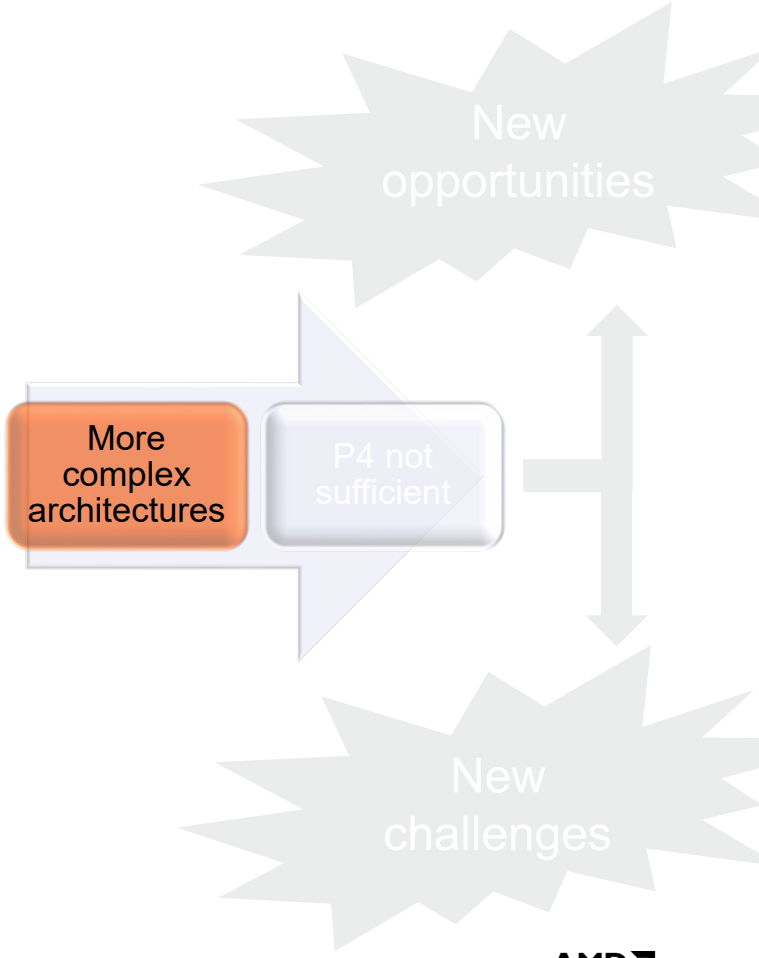
Modifiable table entries

New externs

- E.g., crypto extern; IPsec extern

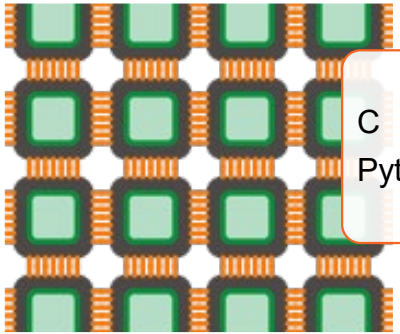
What's Different?





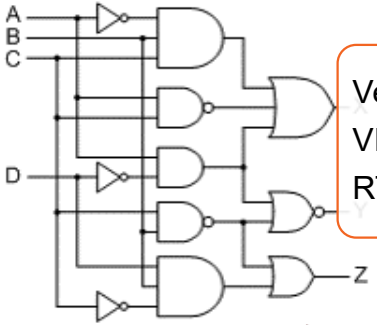
Various Processing Technologies, Different Languages

General purpose processor



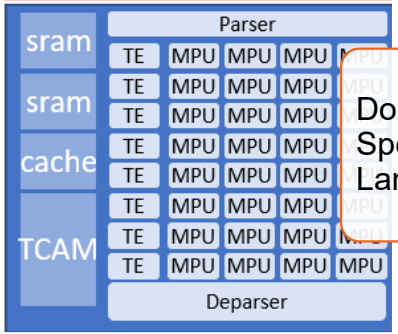
C
Python

FPGA



Verilog
VHDL
RTL

Application specific instruction set processor



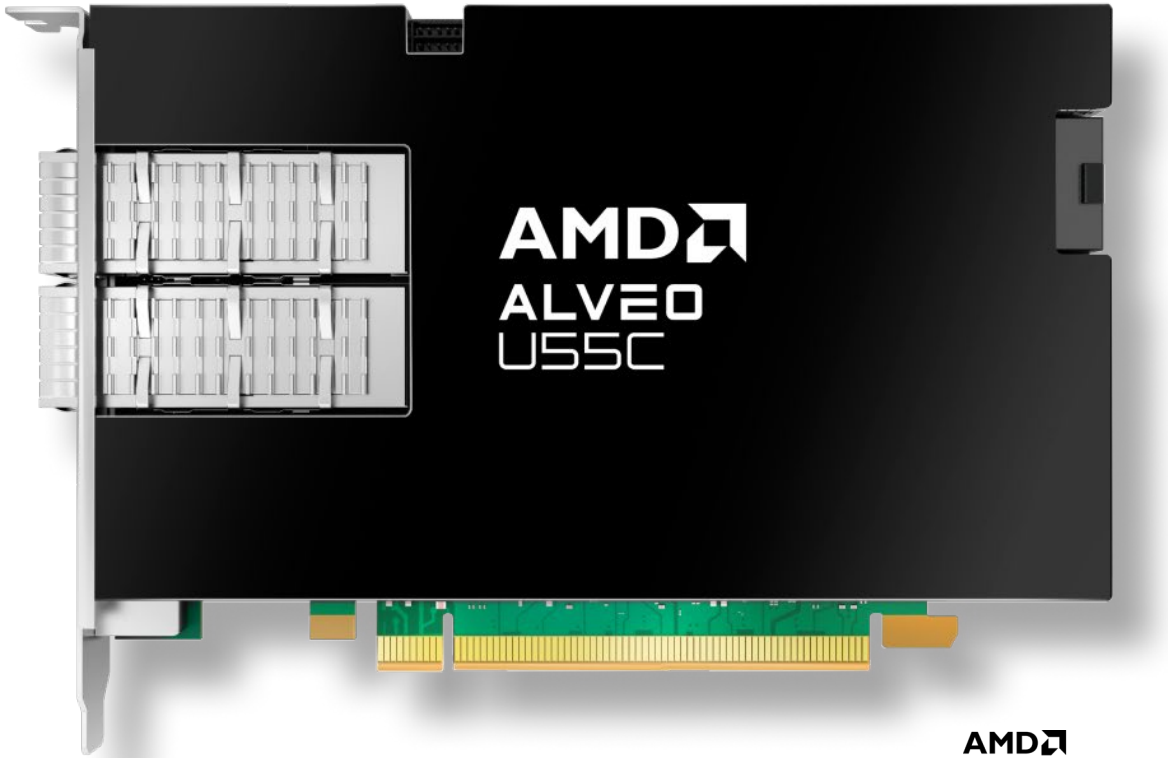
Domain Specific Language

Available SmartNIC Options

Fully Programmable FPGA-based SmartNICs

End user programmed

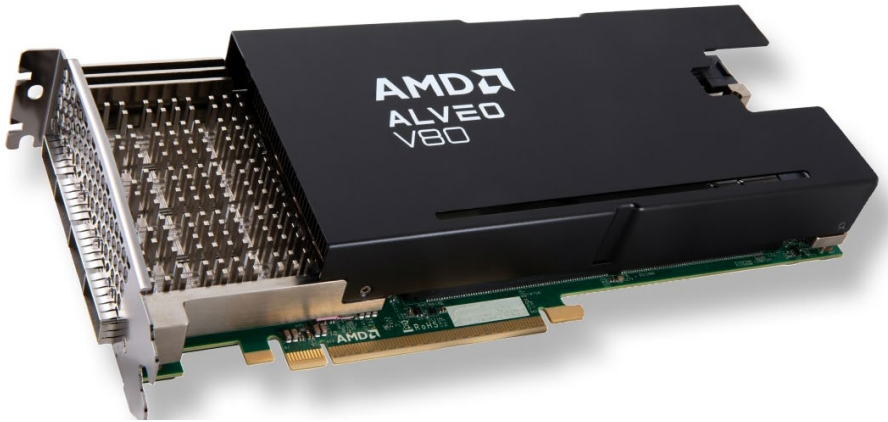
- RTL
- HLS
- P4 to make it simpler

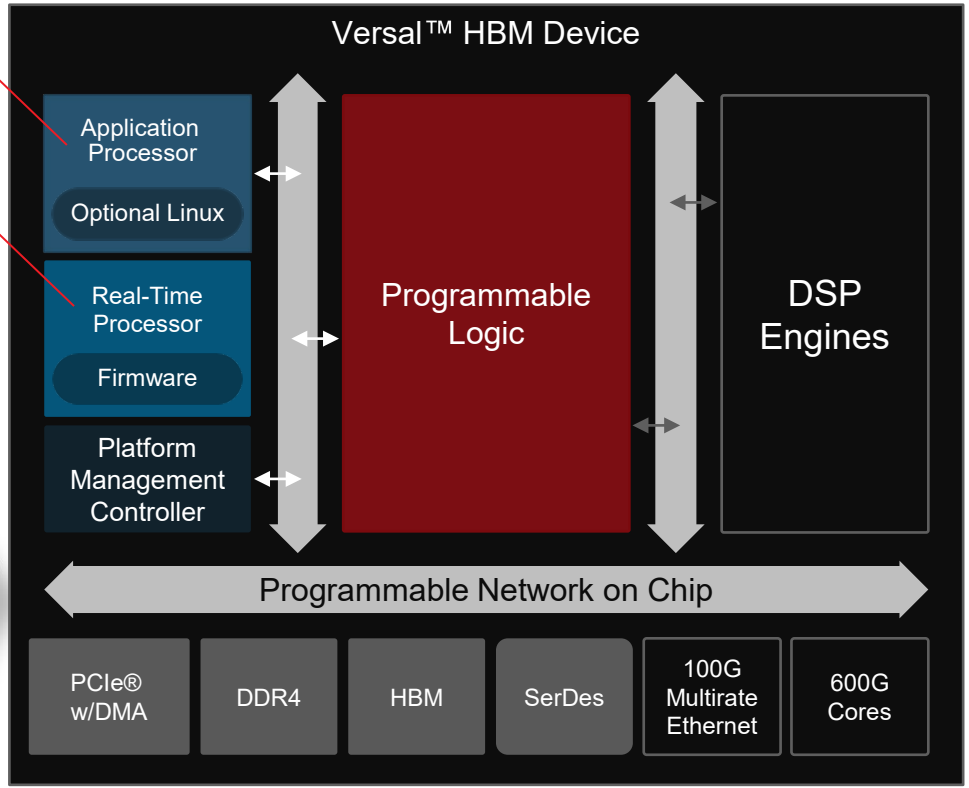


FPGA + arm Cores: Alveo™ V80

Dual-core Arm® Cortex®-A72

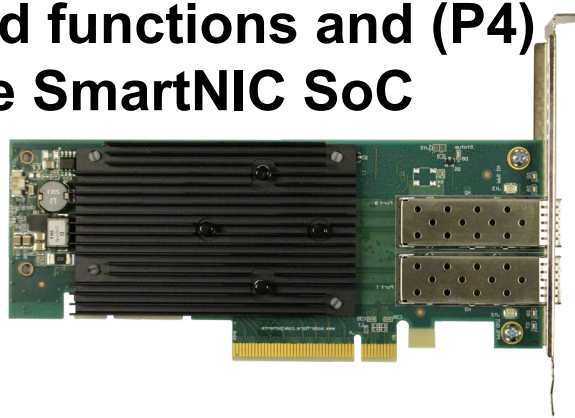
Dual-core Arm® Cortex®-R5F

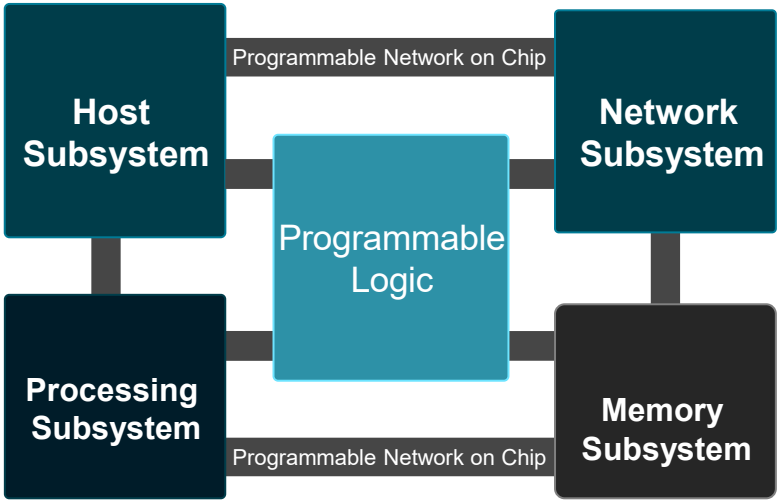




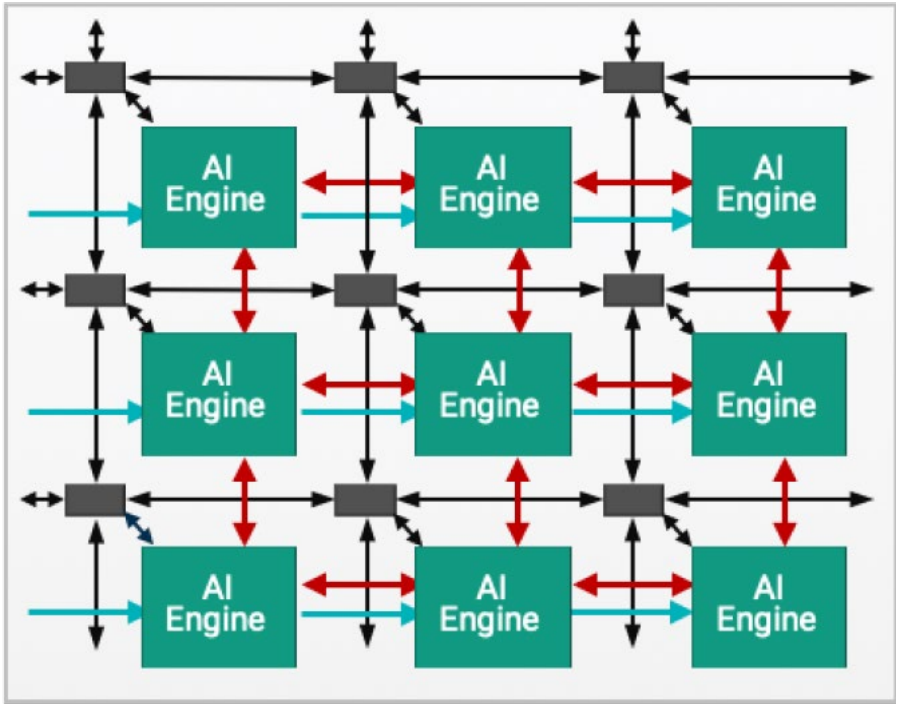
A hybrid approach: fixed functions and (P4) Programmable
The AMD 400G Adaptive SmartNIC SoC

- 2x200 Gbps Ethernet
- 16 lanes PCIe Gen 5
- Hardened logic for
  - Maximum performance
  - Minimum latency
  - Maximum power efficiency
- Large FPGA for custom functionality
- Cloud providers and AI

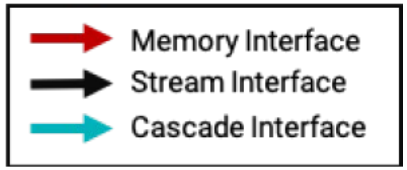




VCK5000: Embedded AI Inference



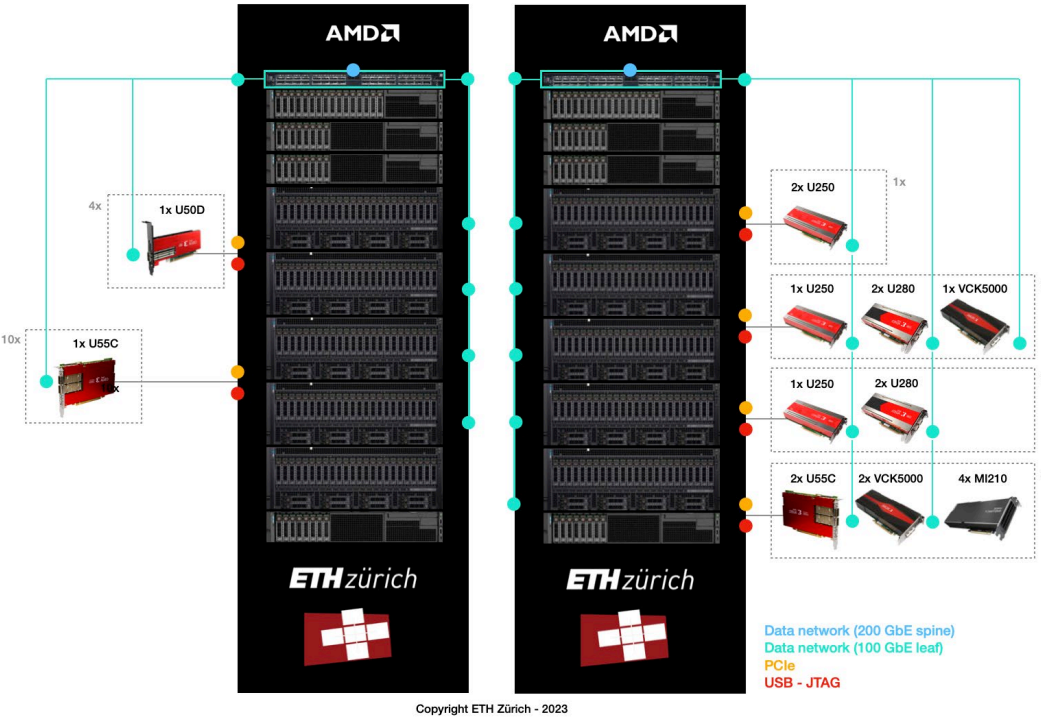
Flexible Interconnect

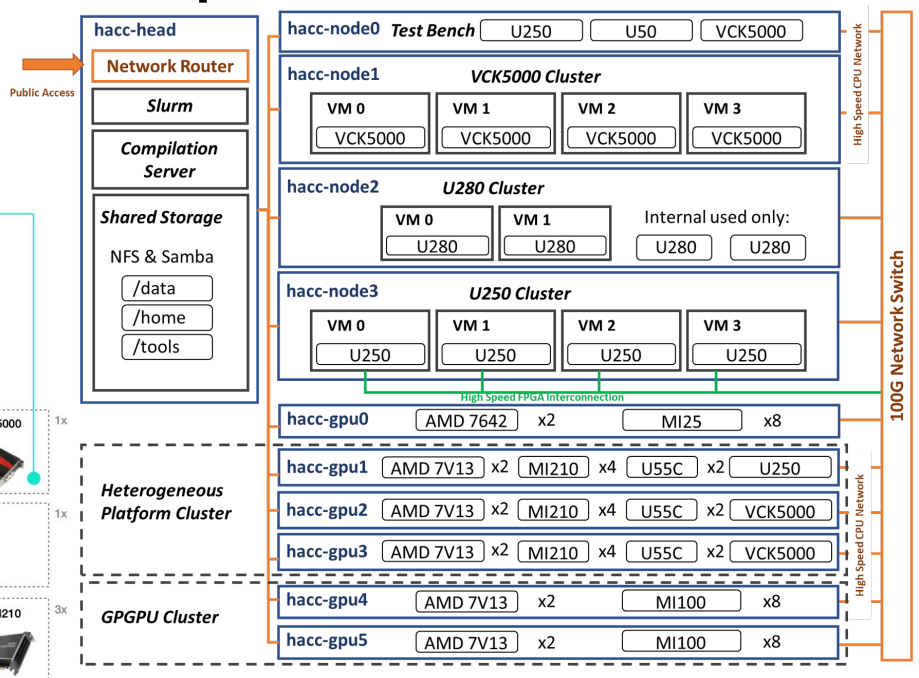


Want to Try?

HACC - Heterogeneous Accelerated Compute Clusters

ETH Zurich





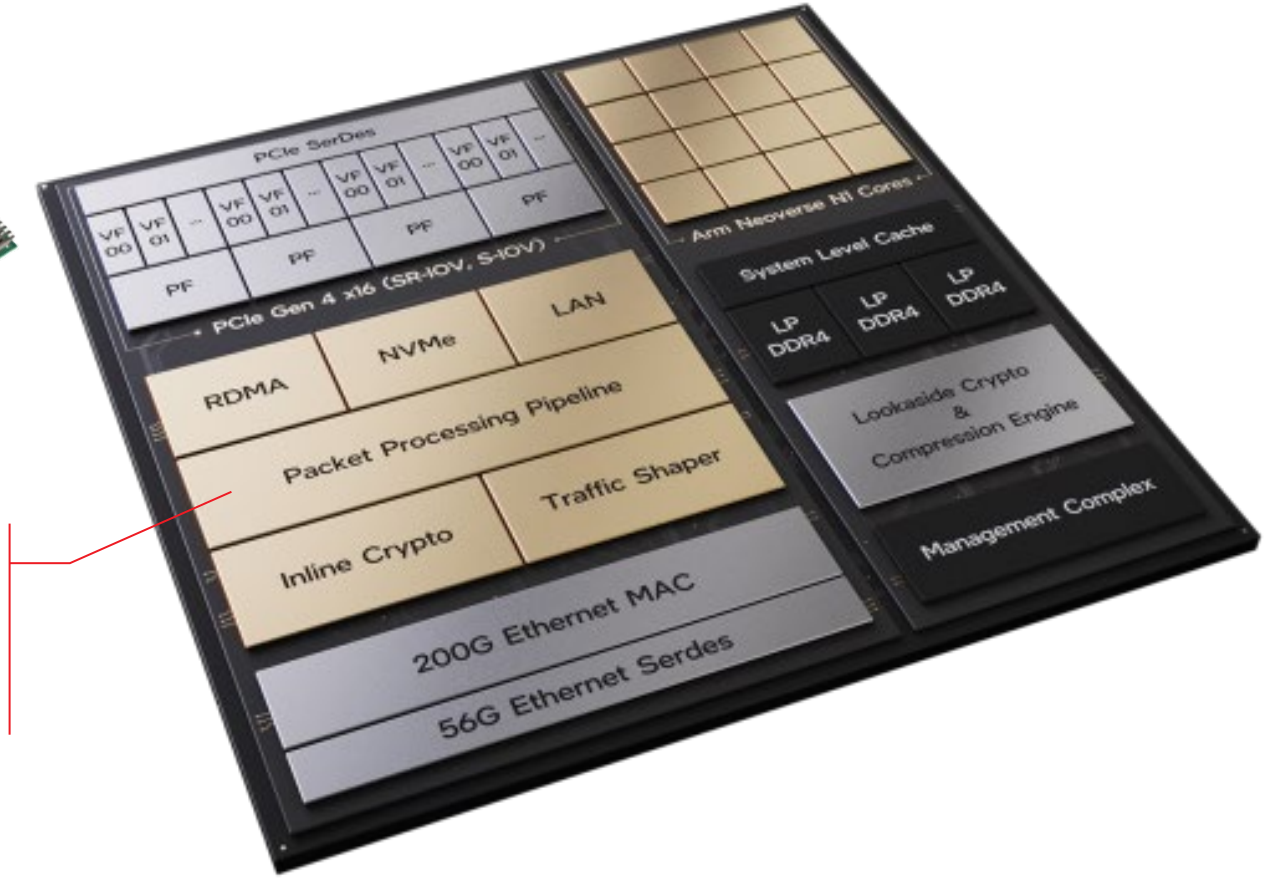
National University of Singapore

https://www.amd-haccs.io/

Architectures including
Application Specific Instruction Set Processors

Intel IPU E2000

Programmable packet pipeline with QoS and telemetry capability



https://www.intel.com/content/www/us/en/products/details/network-io/ipu/e2000-asic.html

NVIDIA Bluefield 3

Host

DRAM

System Level Cache

Arm CPU Complex

PCIe Switch

NIC Subsystem

DPA

RDMA / TCP / UDP

ASAP²

Ethernet / InfiniBand

ConnectX-6

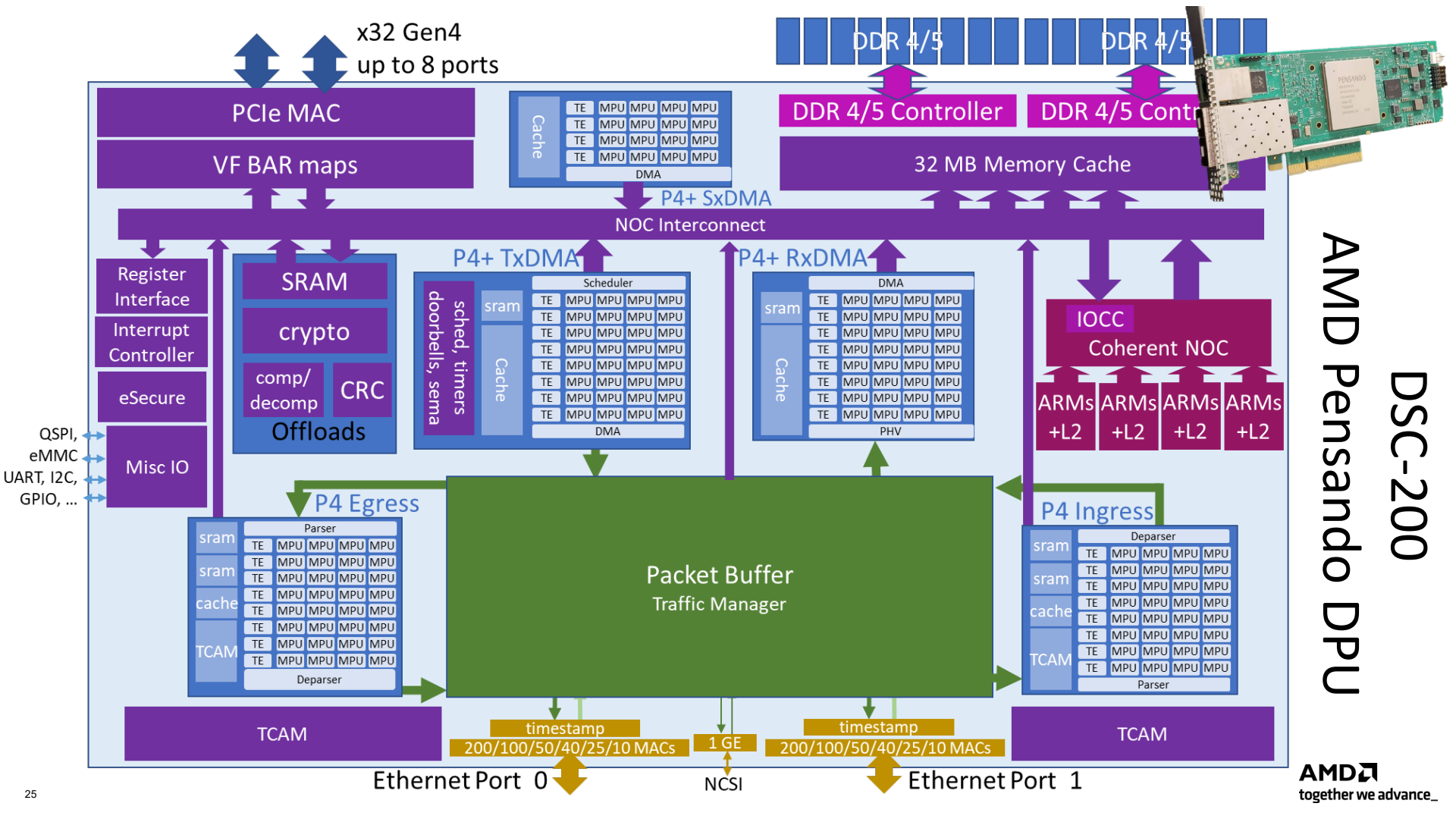
Data Acceleration

PCIe Switch

SSDs / GPU

Network

https://hc33.hotchips.org/assets/program/conference/day1/HC2021.NVIDIA.IdanBurstein.v08.norecording.pdf

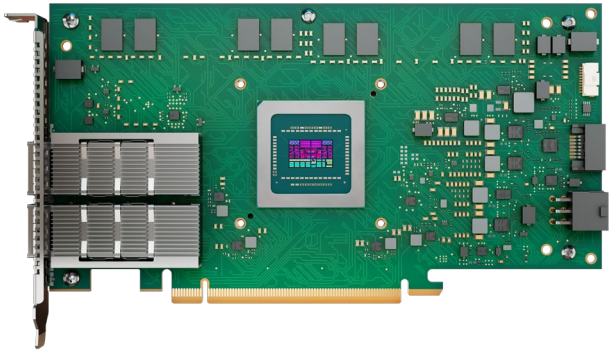


Upcoming

AMD Pensando™ Salina 400

Best DPU for evolving front-end networks

3rd Gen	400G	232 P4 MPU	2x DDR5	16 N1
Software Compatible	PCIe® Gen 5 2x400GE	Multi-Services	102GB/S Memory Bandwidth Up to 128 GB DDR	ARM Cores

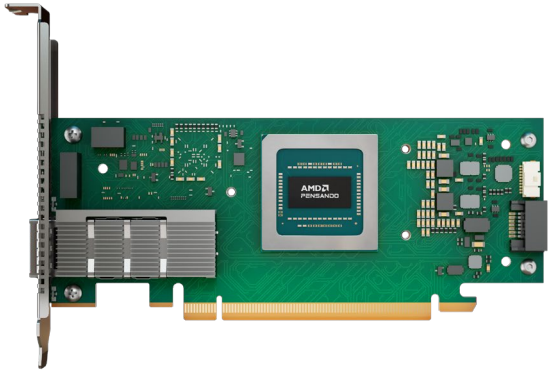


AMD Pensando™
DPU choice for hyperscalers

Upcoming

AMD Pensando™ Pollara 400

Industry's first ultra ethernet consortium ready AI NIC



Programmable Hardware Pipeline

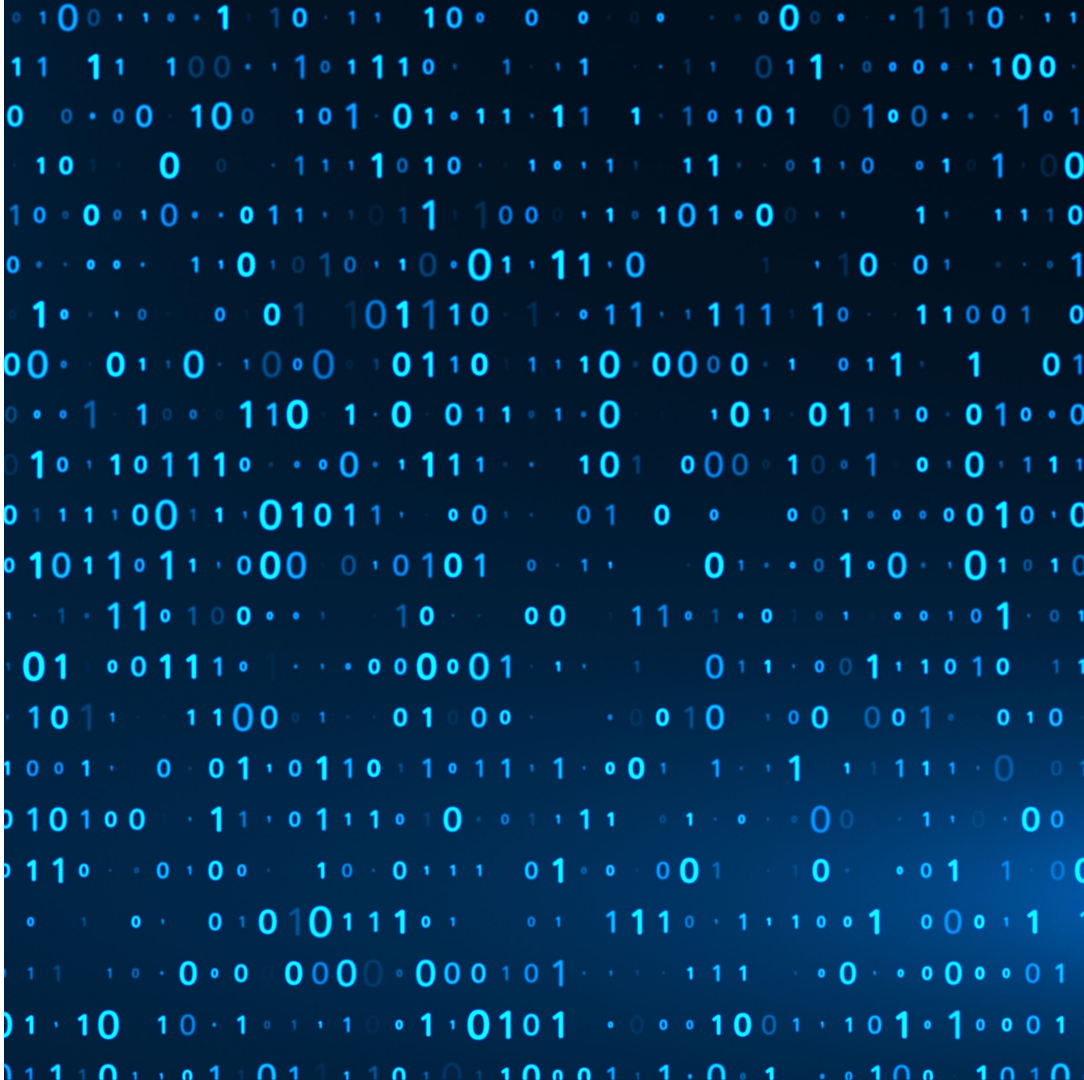
Up to 6x Performance Boost*

400 Gbps

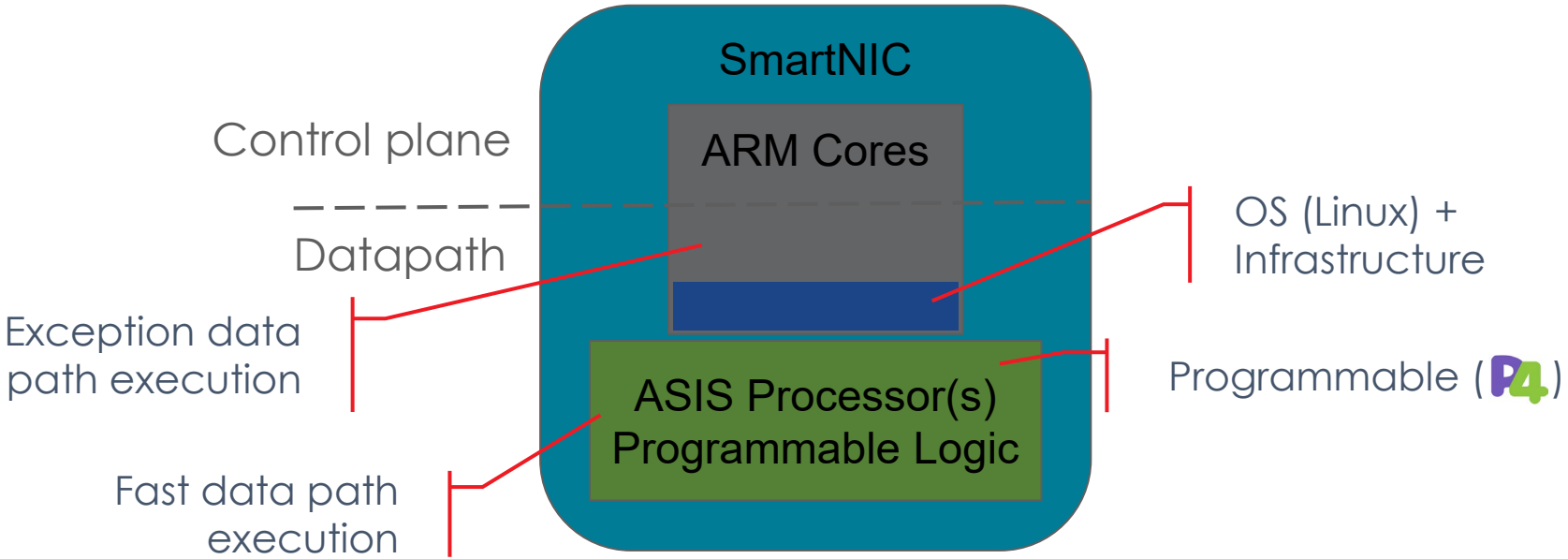
- Open Ecosystem
- UEC Ready RDMA
- Reduction in Job Completion Times
- High Availability

Based on comparison of RDMA vs. RoCEv2. * Reference: STrack: A Reliable Multipath Transport for AI/ML Clusters , July 2024

And The Software?

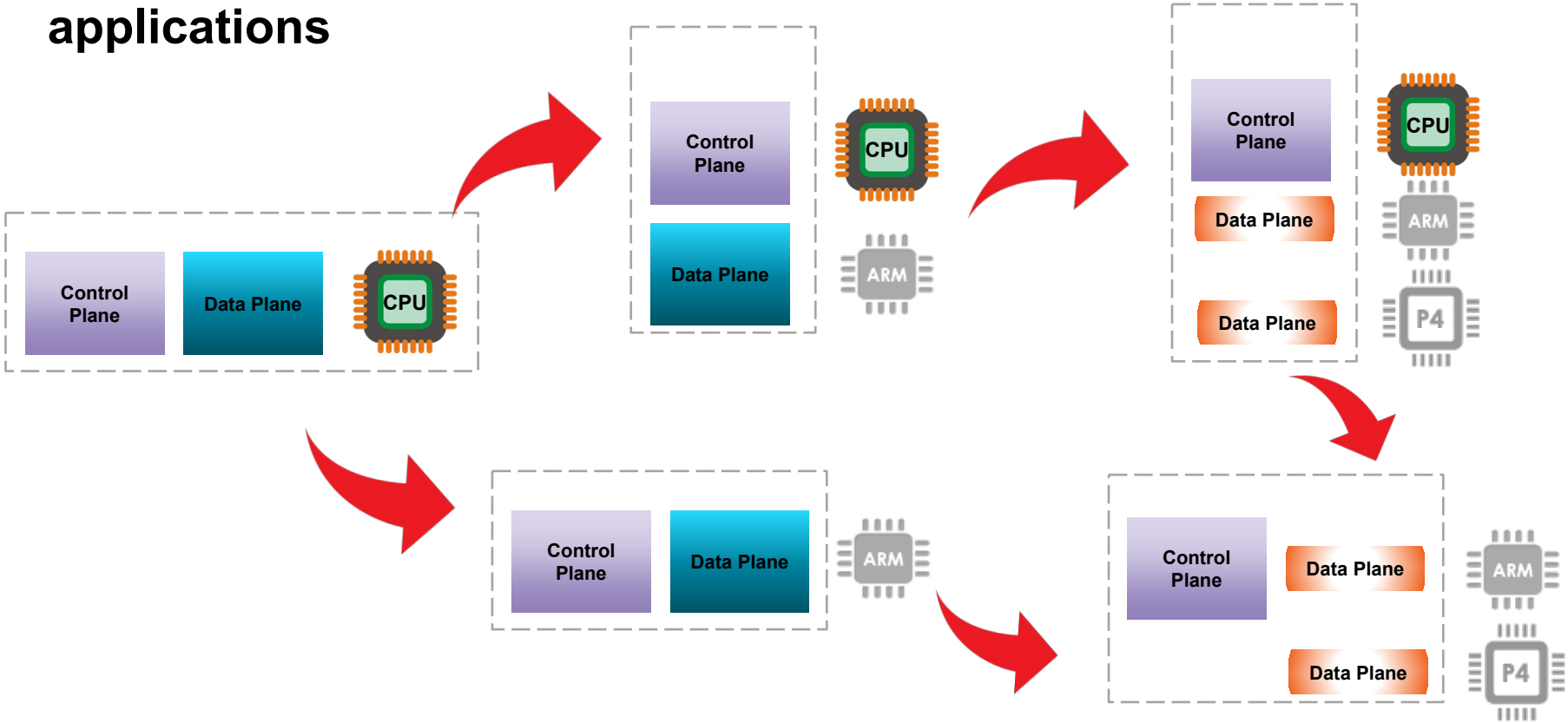


Hardware and Software Architecture at a Very High Level

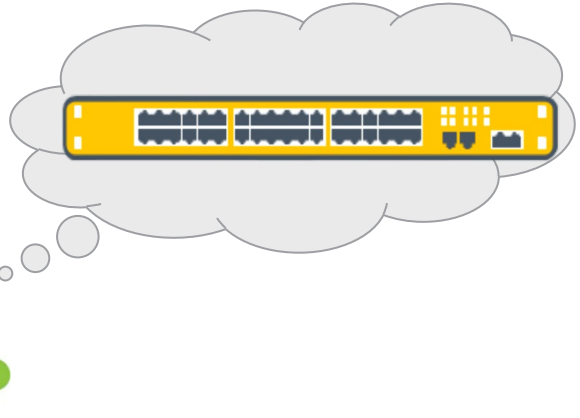


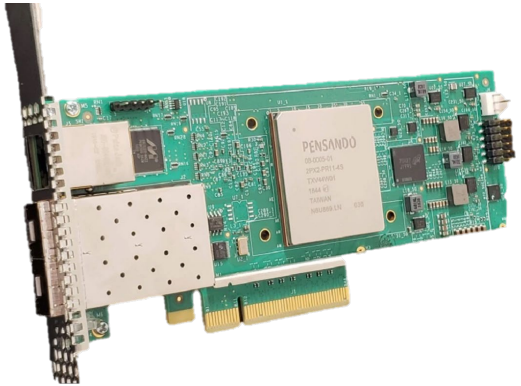
Software development is challenging

An incremental path to accelerating/offloading existing applications

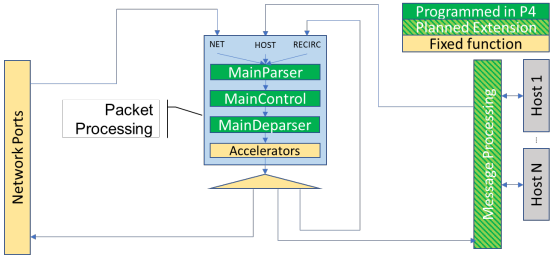


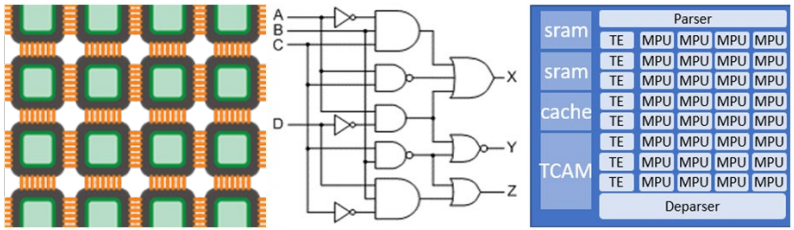
In Summary

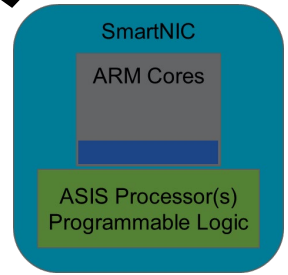




Connected to the host + Lower traffic rate = New opportunities, New challenges







Disclaimer

The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions, and typographical errors. The information contained herein is subject to change and may be rendered inaccurate for many reasons, including but not limited to product and roadmap changes, component and motherboard version changes, new model and/or product releases, product differences between differing manufacturers, software changes, BIOS flashes, firmware upgrades, or the like. Any computer system has risks of security vulnerabilities that cannot be completely prevented or mitigated. AMD assumes no obligation to update or otherwise correct or revise this information. However, AMD reserves the right to revise this information and to make changes from time to time to the content hereof without obligation of AMD to notify any person of such revisions or changes. THIS INFORMATION IS PROVIDED ‘AS IS.” AMD MAKES NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE CONTENTS HEREOF AND ASSUMES NO RESPONSIBILITY FOR ANY INACCURACIES, ERRORS, OR OMISSIONS THAT MAY APPEAR IN THIS INFORMATION. AMD SPECIFICALLY DISCLAIMS ANY IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY, OR FITNESS FOR ANY PARTICULAR PURPOSE. IN NO EVENT WILL AMD BE LIABLE TO ANY PERSON FOR ANY RELIANCE, DIRECT, INDIRECT, SPECIAL, OR OTHER CONSEQUENTIAL DAMAGES ARISING FROM THE USE OF ANY INFORMATION CONTAINED HEREIN, EVEN IF AMD IS EXPRESSLY ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

Third-party content is licensed to you directly by the third party that owns the content and is not licensed to you by AMD. ALL LINKED THIRD-PARTY CONTENT IS PROVIDED “AS IS” WITHOUT A WARRANTY OF ANY KIND. USE OF SUCH THIRD-PARTY CONTENT IS DONE AT YOUR SOLE DISCRETION AND UNDER NO CIRCUMSTANCES WILL AMD BE LIABLE TO YOU FOR ANY THIRD-PARTY CONTENT. YOU ASSUME ALL RISK AND ARE SOLELY RESPONSIBLE FOR ANY DAMAGES THAT MAY ARISE FROM YOUR USE OF THIRD-PARTY CONTENT.


AMD