Smart NICs – Evolution and Future Trends

Raghucharan Boddupalli

and

Nilesh Sable

* Some pictures are borrowed from public domain and are purely for educational purpose.

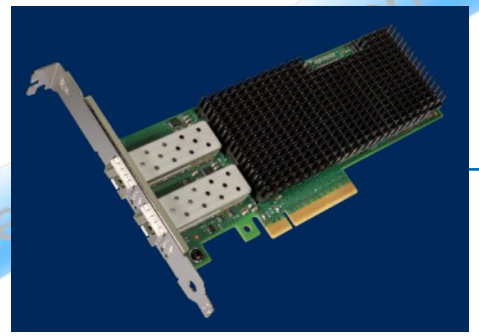
Agenda

- ✓NICs – Foundational to Exotic
- ✓ What are smart NICs ?
- ✓Smart NIC use cases
- ✓Evolution of Smart NICs
- ✓The CPU-DPU-GPU world
- ✓Data center-Cloud evolution
- ✓Generic smart NIC architecture
- ✓The Players in Smart NIC arena
- ✓Acceleration and offload – trajectory
- ✓Use case – Wireless & Telecom
- ✓Use case – Networking
- ✓Use case – Storage
- ✓The battle of ASIC versus FPGA
- ✓AI and the Smarter NICs of future
- ✓Conclusion

NICs – Foundational to Exotic

Foundational	Offload NIC	SmartNIC	DPU	Exotic
Basic interface for network connectivity (popular in the 10/100/1000 and some Nbase-T NICs Less common at 100Gbps+ speeds due to packet processing demands on host CPUs	Offload for common network traffic functions (e.g., TCP/IP stack, limited virtualization features)	Offload functions with additional programmability to offload specific tasks from host systems (e.g., compression/ decompression.) Designed to be a more flexible and expanded offload device	Extended compute, offload, memory, and OS capabilities Designed to be an infrastructure endpoint that exposes resources to the data center and offloads key functionalities for data center scale computing (compute, storage, networking) Higher-levels of compute, offload, memory than SmartNICs	Usually, FPGA-based solutions that have fully customizable pipelines allowing for environment specific optimization Hardware also generally more specific to given deployment scenario

Increasing Cost, Complexity, Capabilities

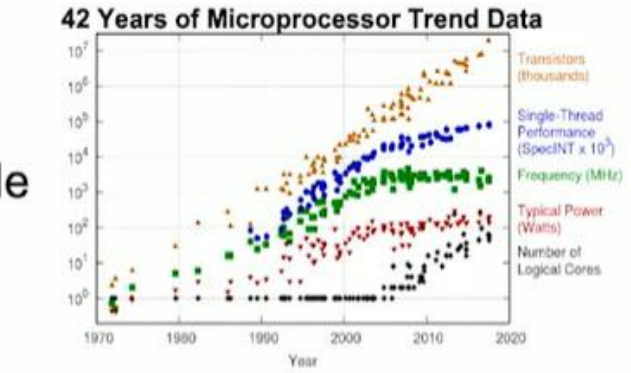


What are smart NICs?

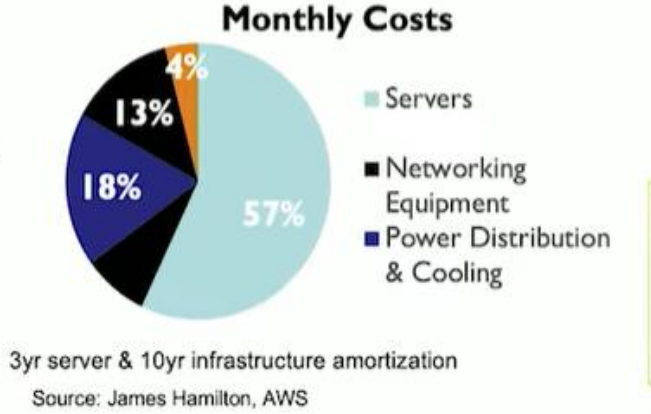
- Network performance is exponentially growing - Increasing popularity of cloud-based computing
- Network speed overpowered CPU performance in the past decade
- Static functionality of a foundational NIC does not comply with evolving Software Defined Network (SDN) policies and Virtual Network Functions (VNF)
- This leads to the development of the SmartNIC - a programmable accelerator that makes data center networking, security and storage efficient and flexible
- SmartNICs offload from server CPUs an expanding array of jobs required to manage modern distributed applications.

Why smart NICs?

- **Moore's Law diminishing returns**
  - Vertical scaling power & cost model no longer viable
- **CPU costs increasing**
  - Economic benefits to limiting core count
- **Multi-socket interconnect bottleneck**
  - I/O, memory transactions across interfaces add latency
  - 2nd socket often used to get more memory and I/O
  - TCO penalty for 2nd socket
- **Distributed cloud architecture**
  - Smaller fault domains



Original data up tot the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010 – 2017 by K. Rupp



3yr server & 10yr infrastructure amortization

Source: James Hamilton, AWS

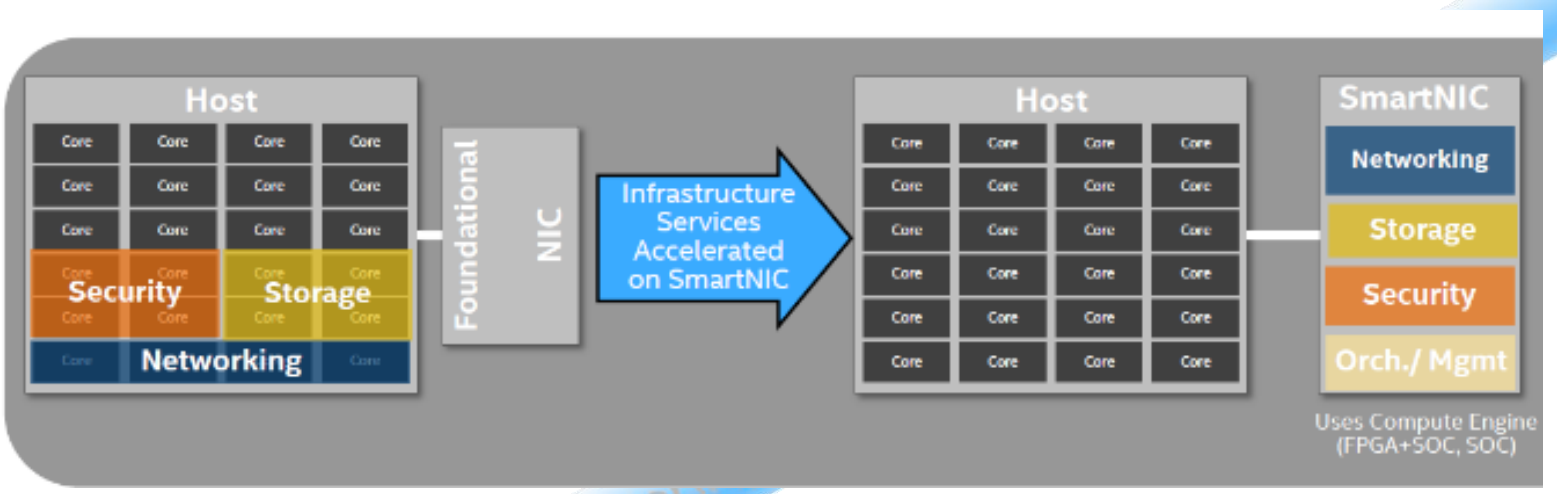
*Courtesy - SNIA

Smart NIC use cases...

- SmartNIC bring computing to the network side, and this makes possible to add more protocols, new virtual functions, offload the network stack, and so on.
- Network virtualization (by offloading VXLAN, NVGRE, or Geneve protocols) or even virtual switches can use SmartNIC, for example, to provide a programmable data path for virtual switch acceleration.
- The following table provides some example of interesting networking functions provided by SmartNICs

SmartNIC function	Use case	Hardware component	Software component
Packet inspection	Intrusion detection, firewall, malware attack, load balancing	Packet marking, filtering, classification, header re-write	Rules definition, control plane
Flow table functions	vRouter, OVS, firewall	packet switching	Switching rules, classification rules and flow tables
Secure networking	Layer2/3/ encryption	Encrypt/ decrypt	Key management
RDMA	Faster bulk data transfer	Transport, networking	Addressing, connections
DPDK/OVS	NFV	packet switching	Rules, reporting
VXLAN overlays	Private/public cloud	Tunneling, encap/decap	rules and framework
NVME-oF	Flash storage	RDMA, protocol layer of NVME-oF	Connection setup, RAID

Smart NIC evolution…



- With use of foundational NICs, Storage, Security and networking services are run on Host CPU
- These services consume as much as 30-35% CPU resources.
- SmartNIC provides multiple flexible and programmable acceleration engines.
- Specialized hardware to handle infrastructure services and packet processing functions

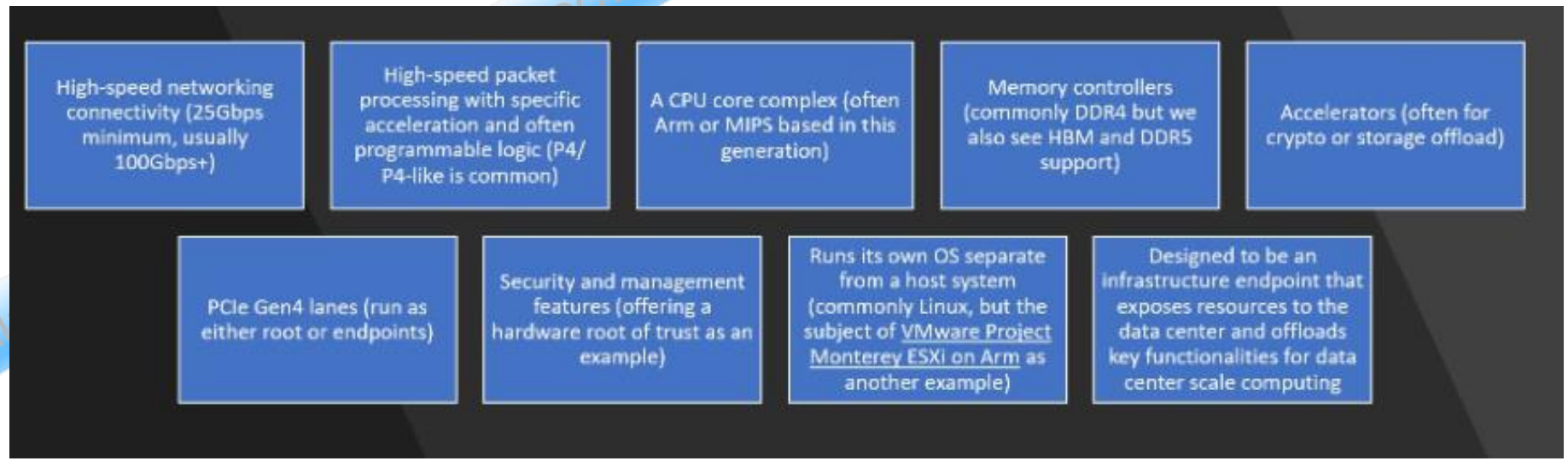
The SmartNIC - IPU

- Highly programmable packet processing engine, NVM Express storage interface scaled up from Intel® Optane™ Technology, next generation reliable transport, advanced crypto, and compression acceleration
- Open-source software design: Infrastructure Programmer Development Kit (IPDK) leverages and extends Data Plane Development Kit (DPDK) and Storage Performance Development Kit (SPDK). IPDK is vendor agnostic and runs on a CPU, IPU, DPU, or switch.

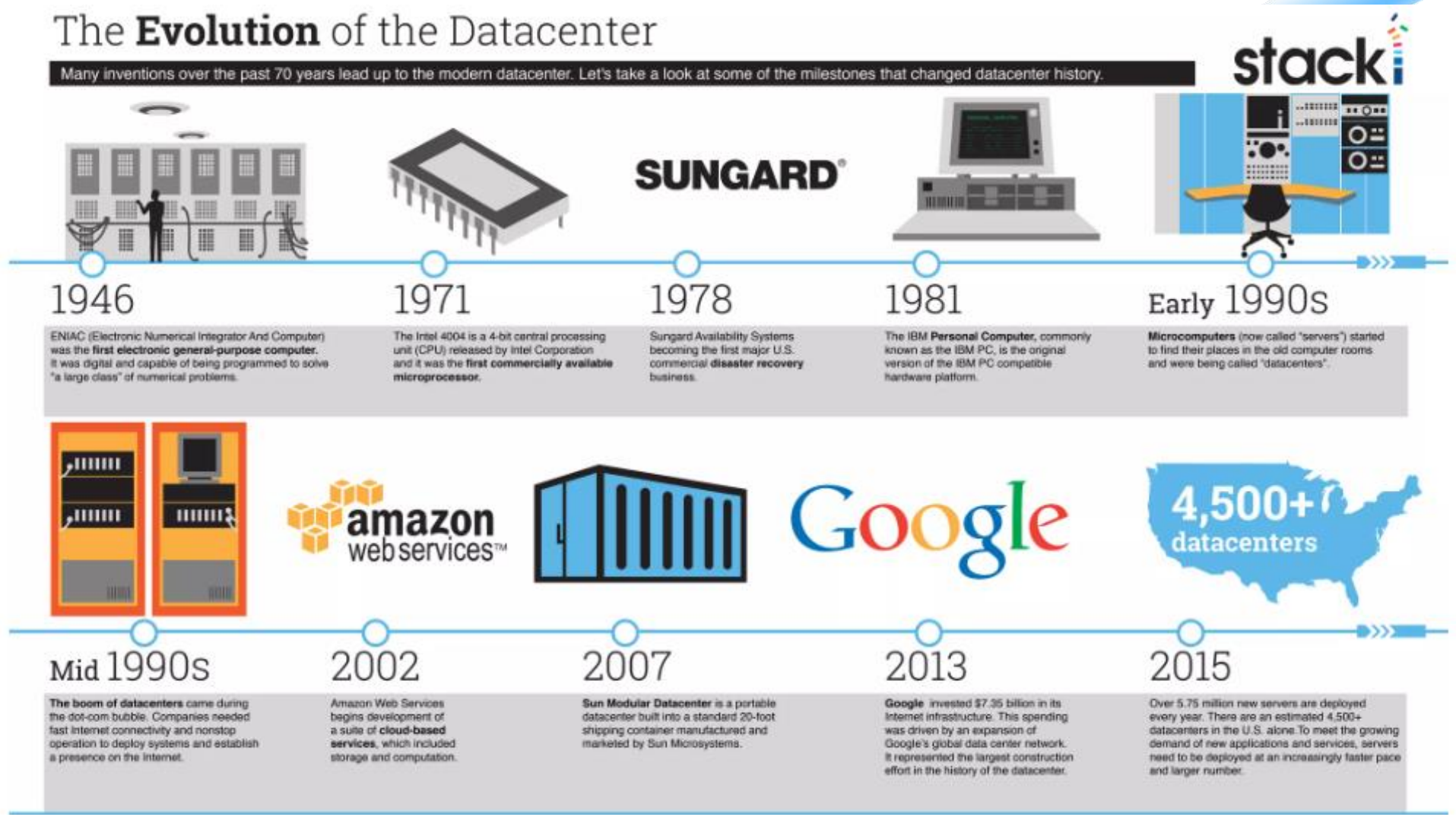
Products	Features	Target Acceleration Workloads
Intel® IPU E2000 	• 2 x 100 GbE or 1 x 200 GbE connectivity • Up to 16 Arm Neoverse N1 Cores • PCIe 4.0 x16 • Up to 48GB DRAM	• Packet processing • OVS • NVMeOF and Storage • RDMA/RoCEV2 • Traffic shaping and QoS • Security: Inline and Lookaside Crypto with Compression
Intel® IPU Platform F2000X-PL 	• 2 x 100 GbE connectivity • Intel® Agilex-F FPGA • Intel® Xeon D-1736 Processor • 32GB DRAM	• Packet processing • OVS • NVMe-oF • Security/Isolation • Crypto • RDMA/RoCEV2
Intel® IPU Platform C5000X-PL 	• 2 x 25 GbE connectivity • Intel® Stratix® 10 DX FPGA • Intel® Xeon D-1612 Processor • 20GB DRAM	• Packet processing • OVS • RDMA/RoCEV2

The SmartNIC - DPU

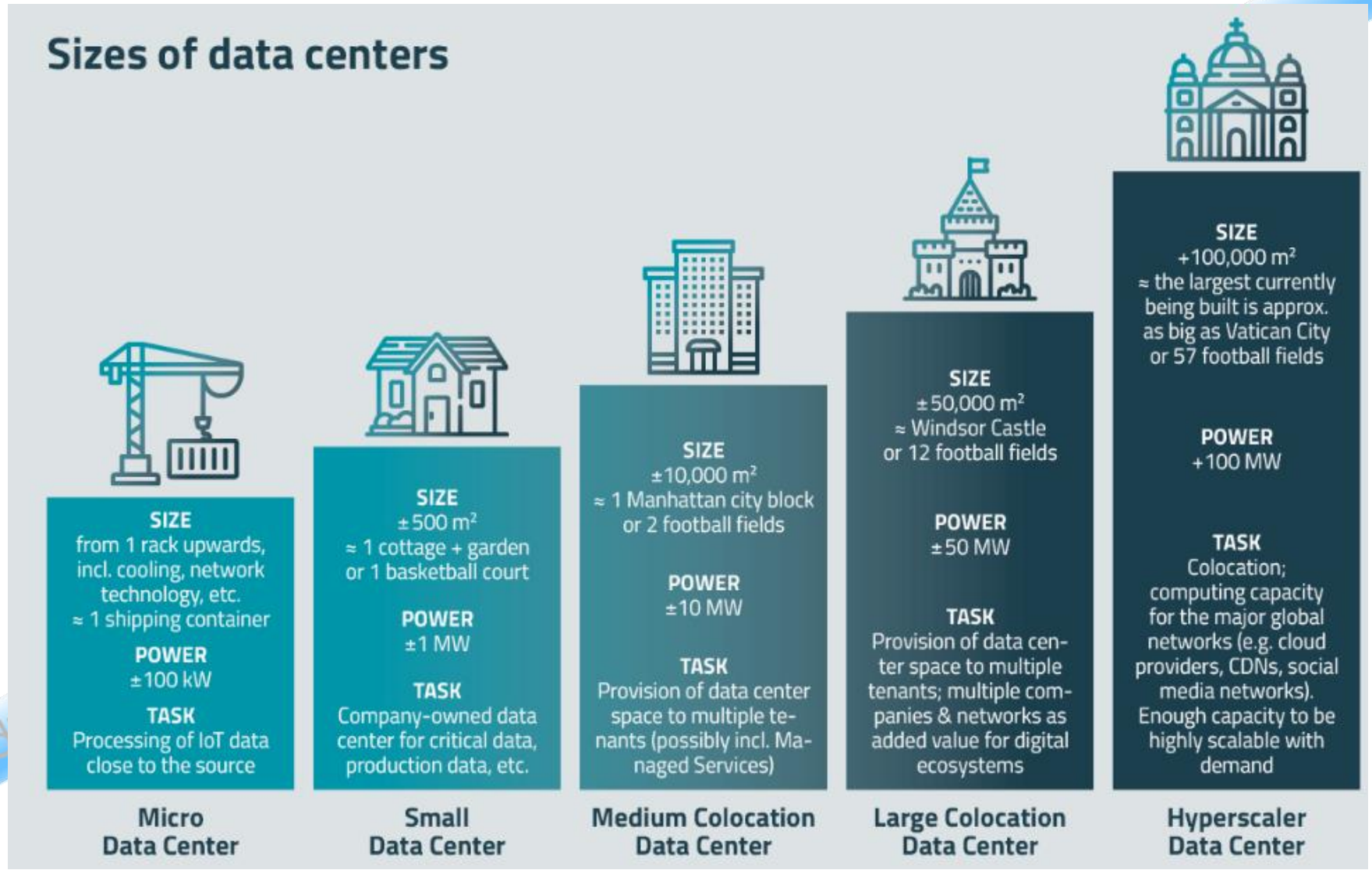
- DPUs are a major evolution of the SmartNIC. They include the offload, flexible programmable pipeline, processing and CPU of SmartNICs.
- DPUs include custom chips and, in some cases, customized field-programmable gate arrays or custom application-specific integrated circuits.
- A DPU can support much more than a SmartNIC, including networking based on P4 programmable pipelines, stateful Layer 4 firewalls, L2/L3 networking, L4 load balancing, storage routing, storage analytics and VPNs.
- DPU functionality varies by vendor. Some of the major players in the market in 2022 are Fungible, AMD Pensando and Marvell.



Data center-Cloud evolution



Data center evolution contd...



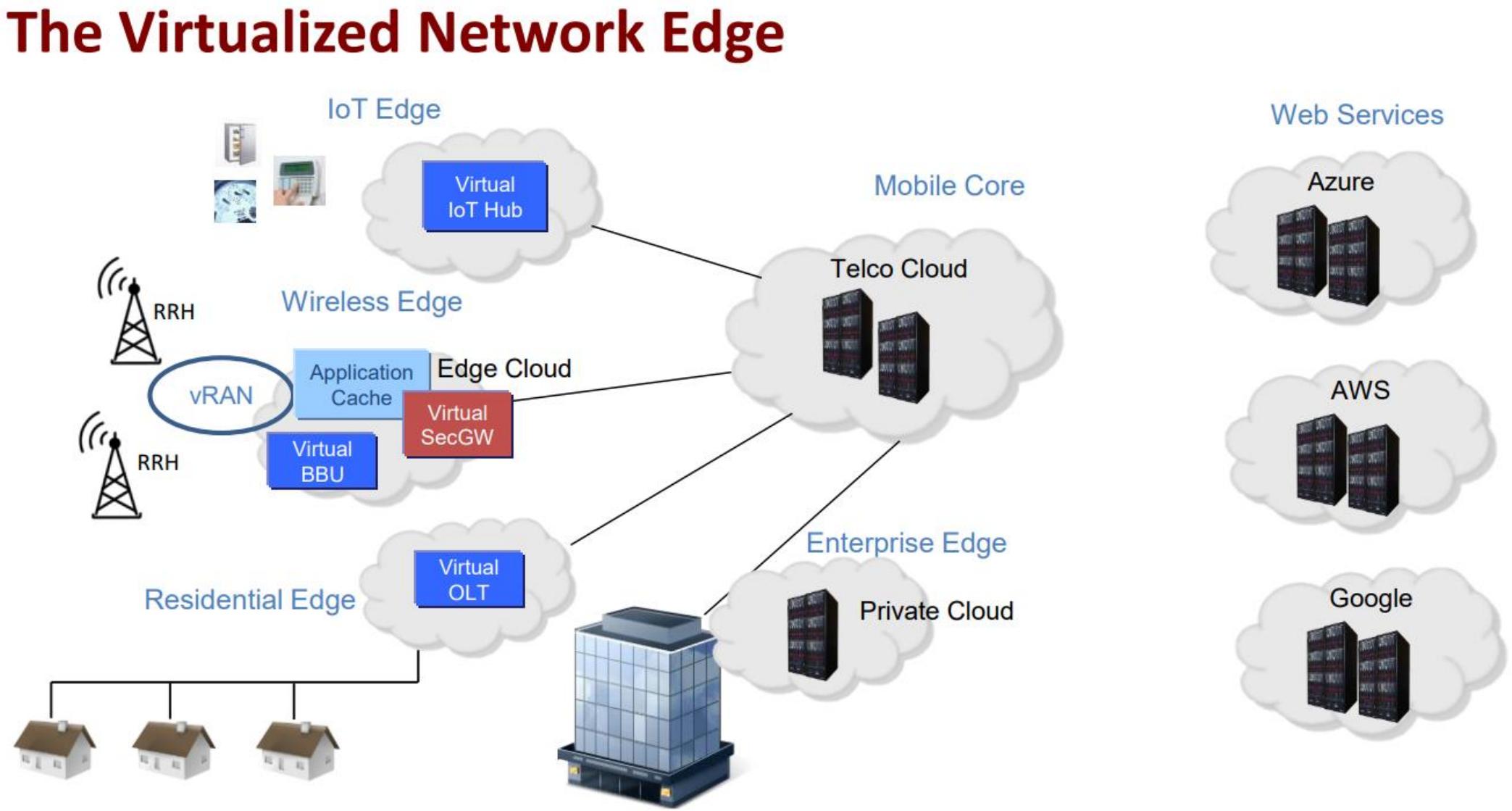
Smart NIC players...

- Achronix : Speedster7t FPGA series- FPGA
- AMD(Xilinx;Pensando) Alveo series、Elba、Capri – FPGA, SoC
- Asterfusion: Helium SmartNIC– SoC ARM+ASIC+ Dedicated Accelerator
- AWS:Nitro system – SOC
- Azure : Catapult – GP SoC
- Broadcom: Stingray – SoC:ARM+ASIC
- Intel ： IPU- FPGA+X86 SoC、 FPGA+ARM SoC
- Fungible :FI DPU – NP SoC
- Kalray :K200/K200-LP – MPPA DPU Processor
- Marvell : OCTEON 10 DPU – SoC:ARM+ASIC
- Napatech : NT200A02 SmartNIC – FPGA
- Netronome : Agilio Series
- Nvidia: BlueField DPU – SoC:ARM+ASIC+Dedicated Accelerator/GPU Dedicated Accelerator
- Silicom :N5010 , N5110A, P425G2SNxIAONIC – FPGA

Acceleration and offload – trajectory

- It started with Layer 2 protocols offloading
- It has taken a long trajectory to support L4 protocol offloading
- Presence of hardened Crypto Engines enables offloading of security protocols (SSL/TLS)
- With increasing processing power, SmartNICs provide fine grained load balancing of flows
- SmartNICs capable of packet header processing enable packet hints which are useful for latency sensitive applications
- Further, packet characteristic can be analyzed with Deep Packet Inspection which accelerate many use cases

Use case – Wireless and Edge

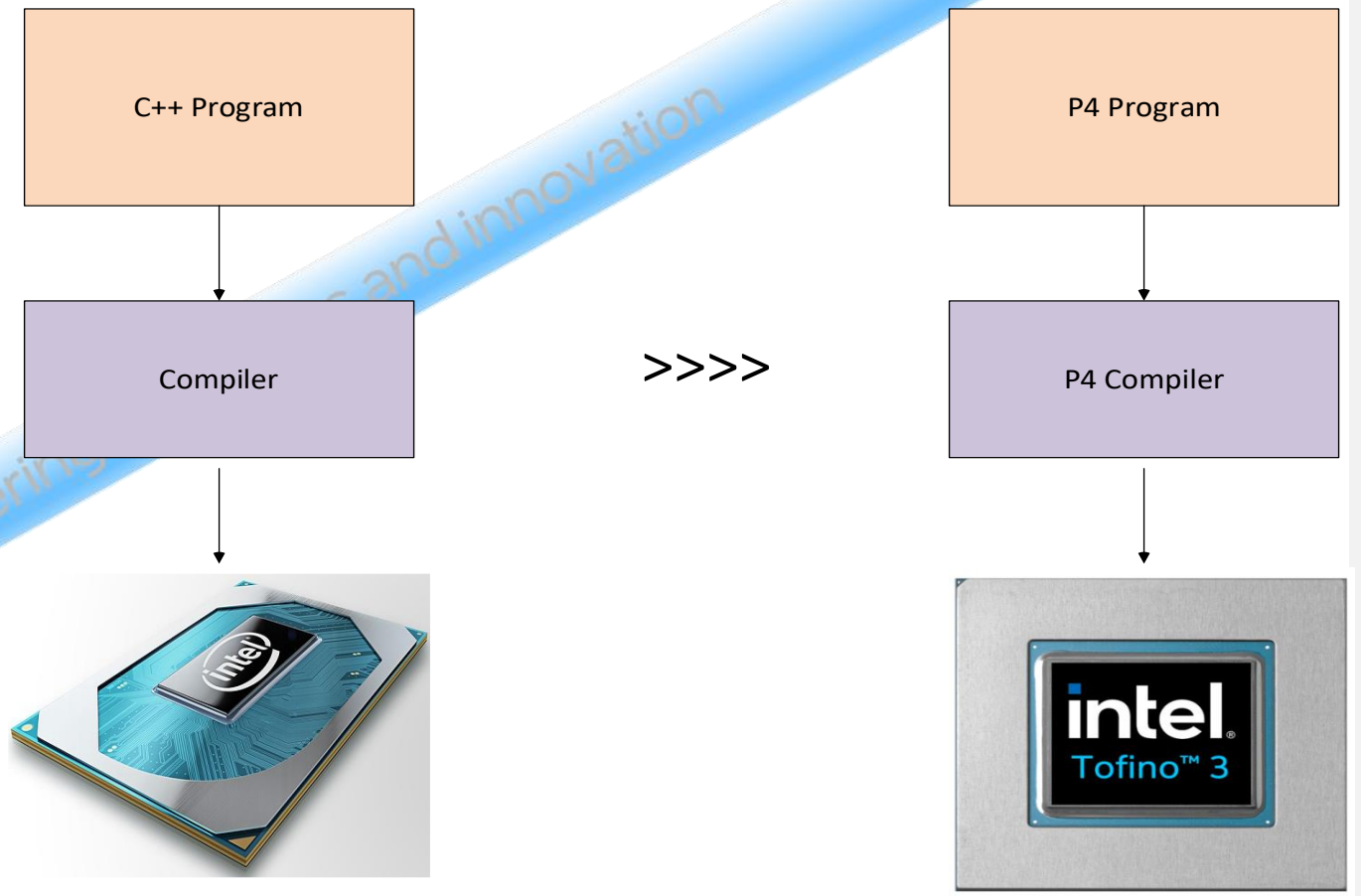


Use case – Networking …

- Data packet processing is much faster at SmartNIC pipeline
- Accelerator chips speed up common processing tasks significantly
  - Packet capture
  - Network management
  - Telemetry
  - Intrusion detection
  - Data decompression and deduplication
  - Routing, Load balancing
  - Firewalling

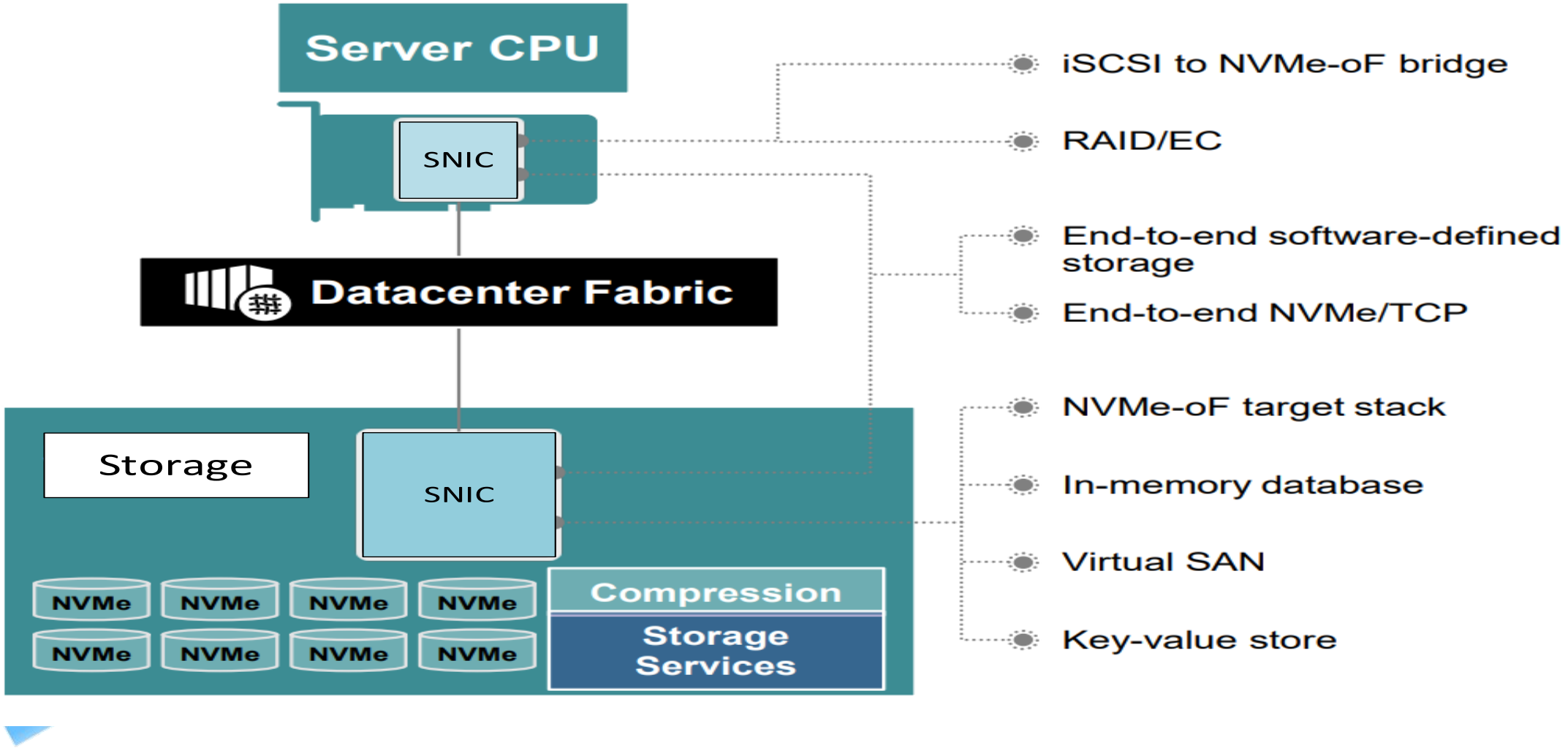
Use case – Networking - Software Defined Networking

- Separation of data plane and control plane
- Shorter time to market as product features are software defined and new protocols can be quickly implemented
- Longer product life as product functionality can be upgraded by modifying software

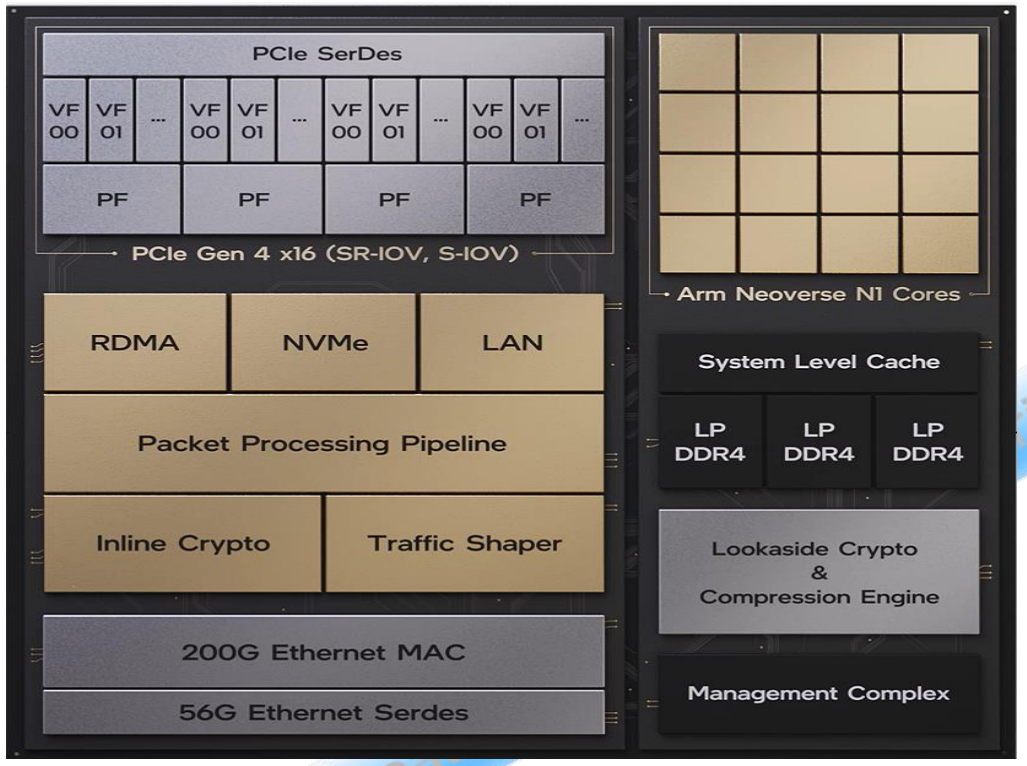


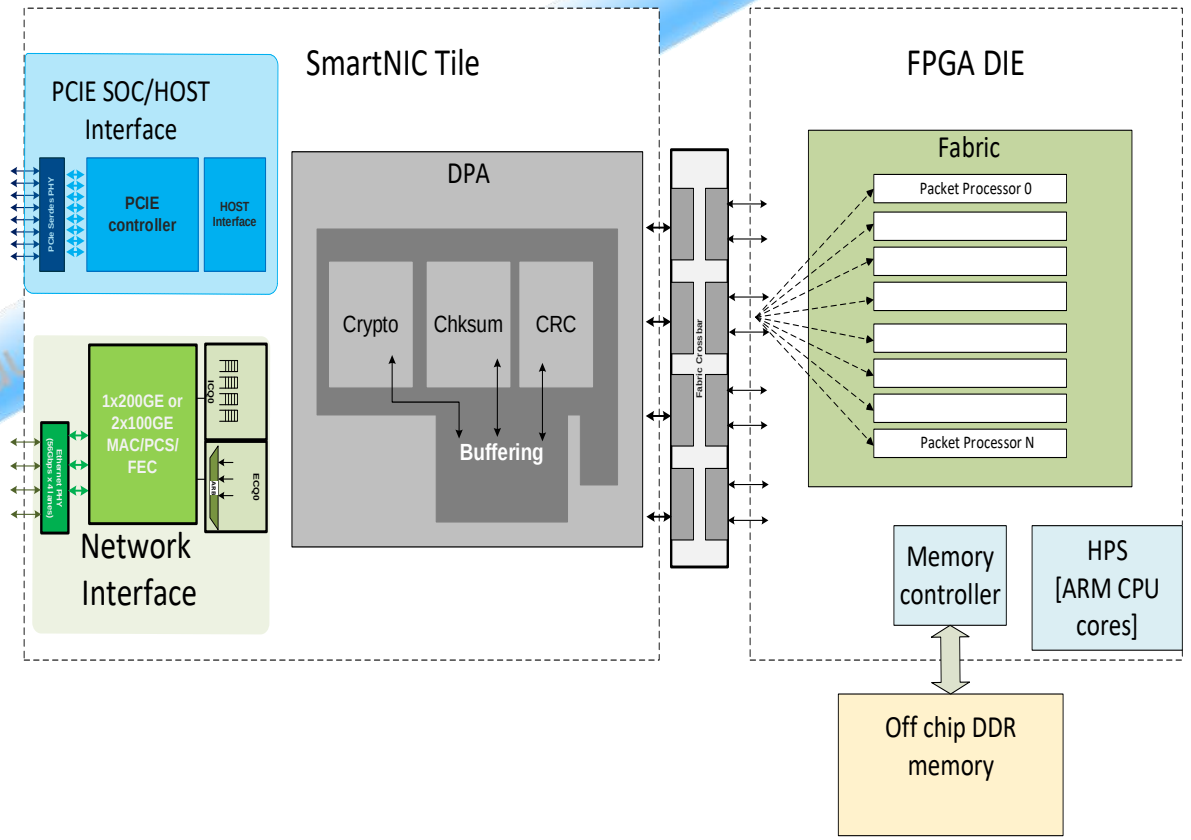
P4: Programming Protocol-Independent Packet Processors

Use case – Storage

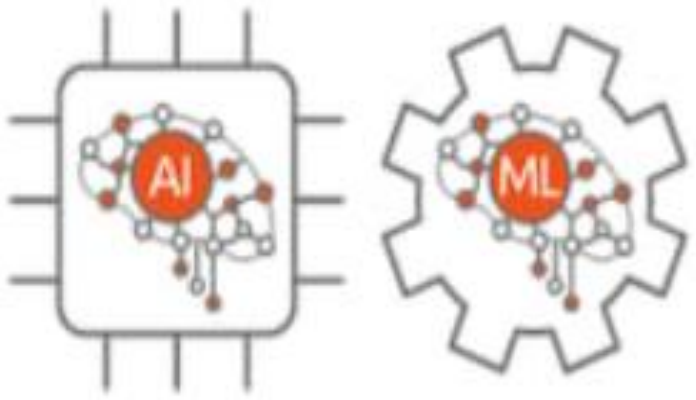


The battle of ASIC versus FPGA SmartNICs





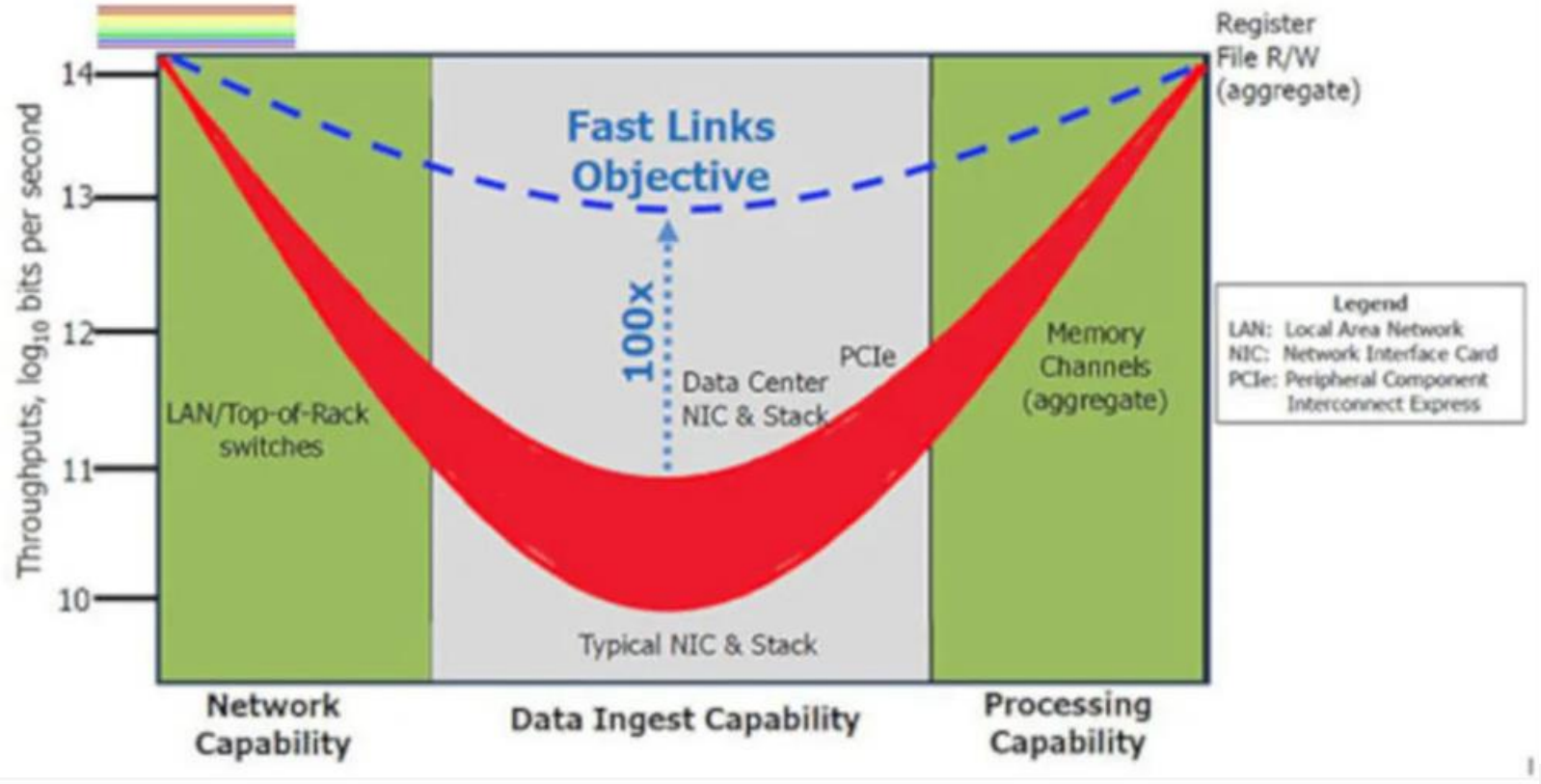
AI and smarter NICs of future



- AI processing for intelligent feature extraction
- Improved handling of traffic flows using AI learning
- Adaptive routing and security applications can receive locally made decisions
- SmartNICs can reduce the total power consumption, delivering AI at scale at lower cost.

FastNICs for AI

- DARPA: Goal is 100x faster network card for tomorrow's AI
- The lack of significantly faster NICs is also becoming a challenge for AI in deep neural network training and image classification.



Conclusion

- SmartNIC technology is a key enabler in the next generation converged architecture by accelerating data processing and data movement
- Data centers become more scalable, secure and cost-effective as tenant applications are not affected by infrastructure workload
- Freed up main CPU cores drive revenue growth
- SmartNICs are rapidly evolving and enabling use cases for DC, Telecom, Edge
- AI is boosting utility of SmartNICs and SmartNICs will in turn boost AI

Q&A

Thank You

https://www.intel.in/content/www/in/en/silicon-innovations/6-pillars/process.html