UNLOCKING GENERAL PURPOSES PROCESSORS TO BOOST PACKET PROCESSING PERFORMANCE

Liang Cunming
Platform Solution Architect
Data Center / Network Platforms Group

Legal Notices & Disclaimers

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Check with your system manufacturer or retailer or learn more at intel.com.

Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. Consult other sources of information to evaluate performance as you consider your purchase. For more complete information about performance and benchmark results, visit http://www.intel.com/performance.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit http://www.intel.com/performance.

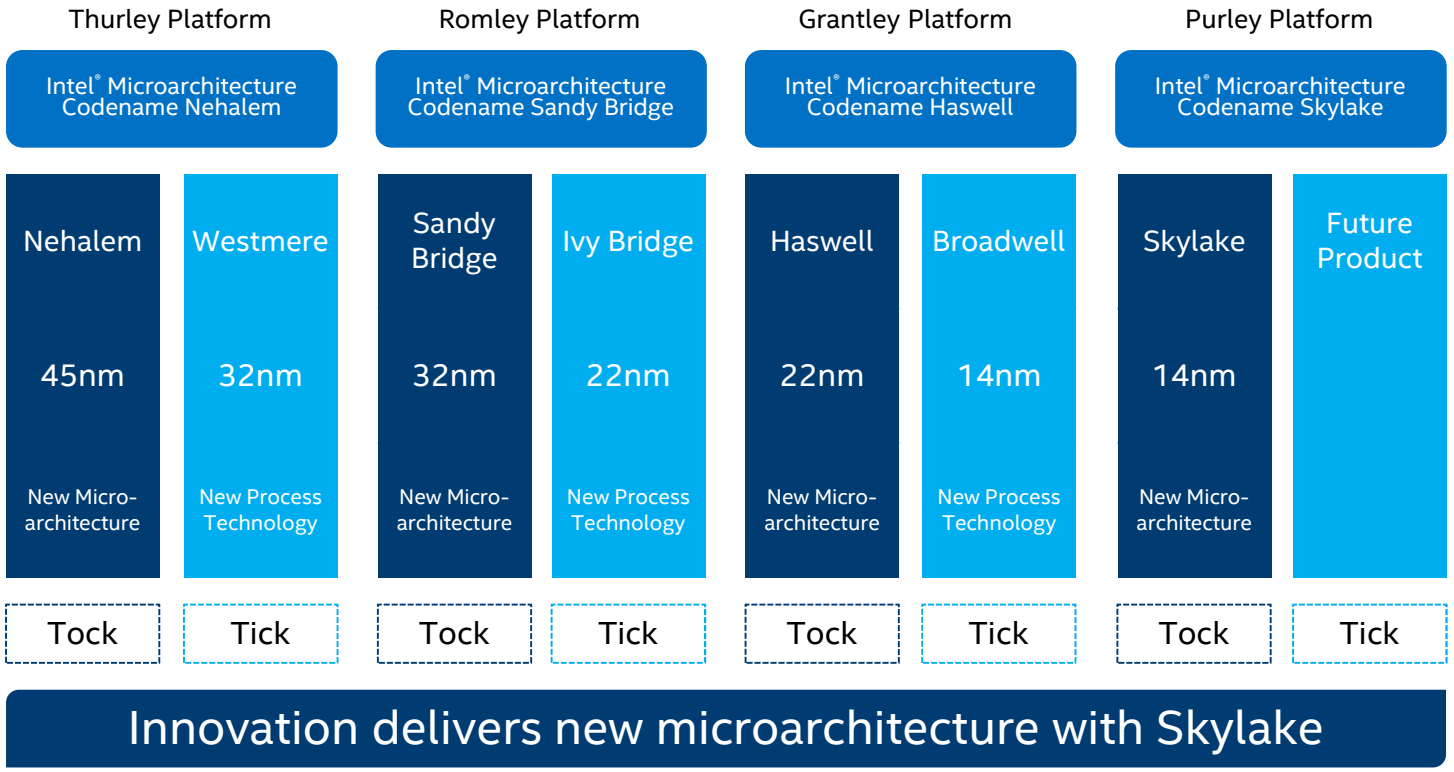
Cost reduction scenarios described are intended as examples of how a given Intel-based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction.

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

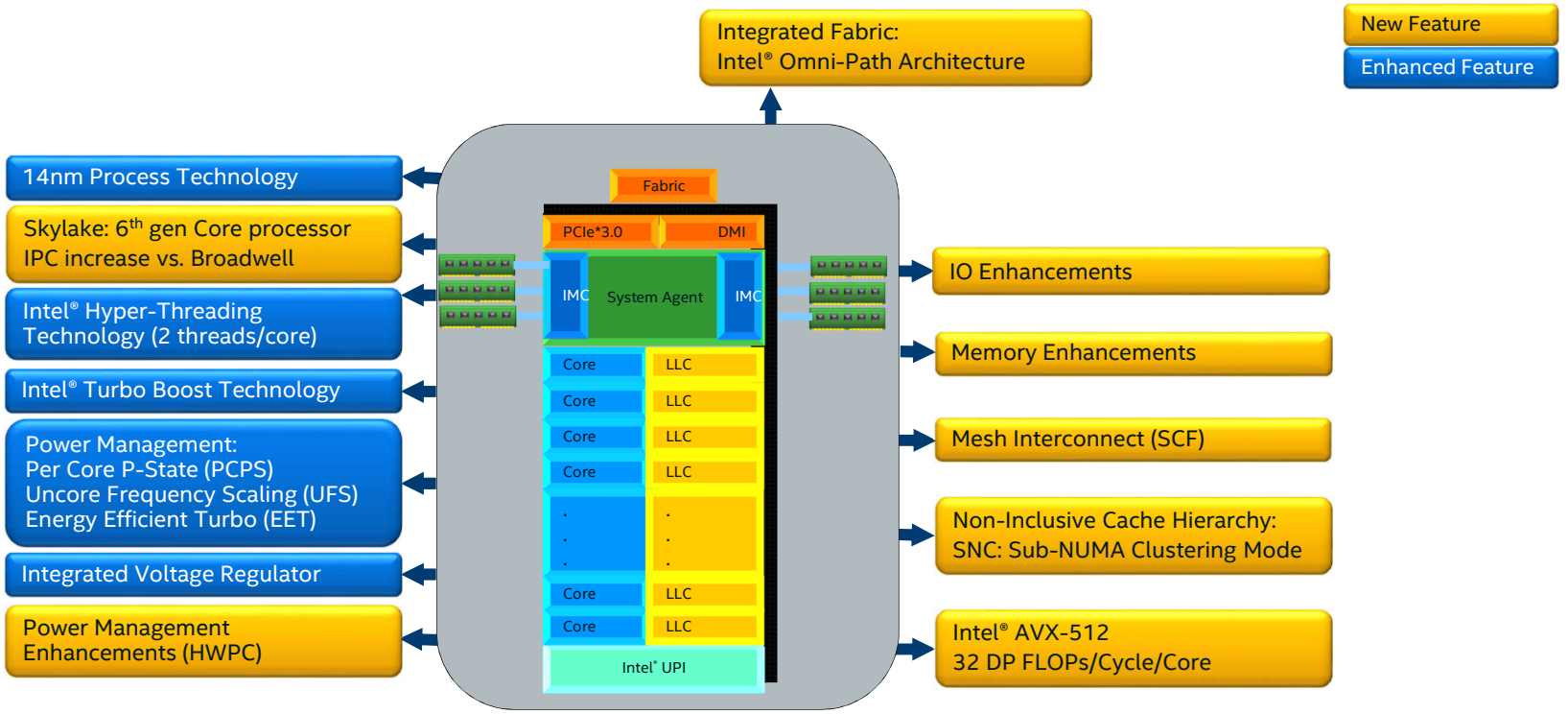
Intel does not control or audit third-party benchmark data or the web sites referenced in this document. You should visit the referenced web site and confirm whether referenced data are accurate.

Tick-Tock Development Model:

Sustained Microprocessor Innovation Leadership

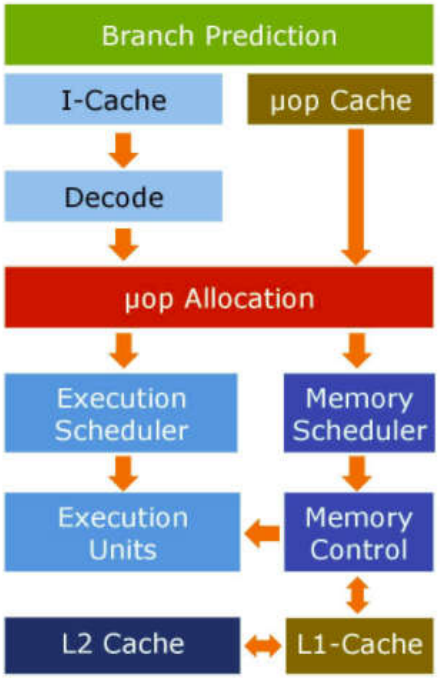


Skylake-SP Server CPU Overview



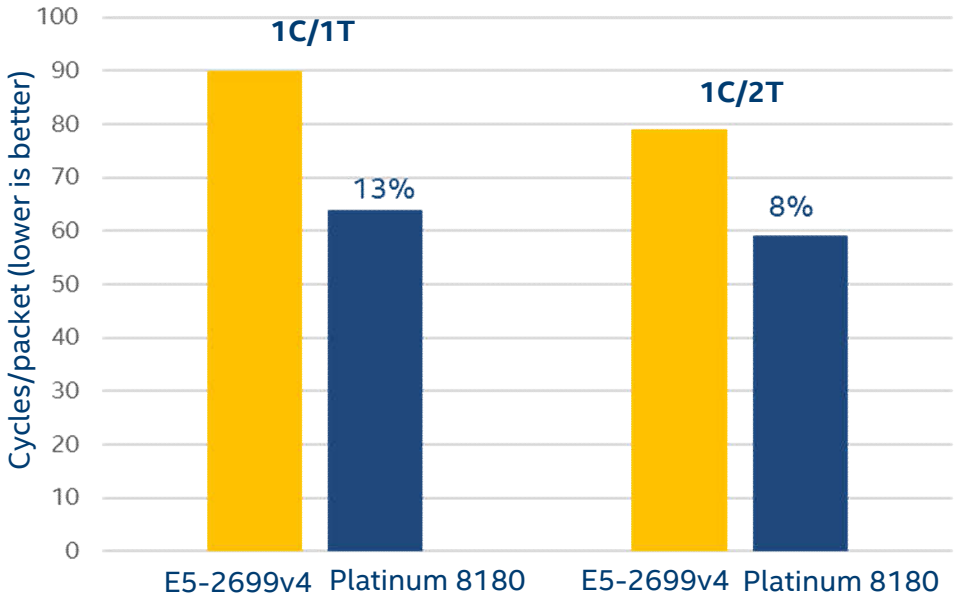
Skylake Core Micro-Architecture

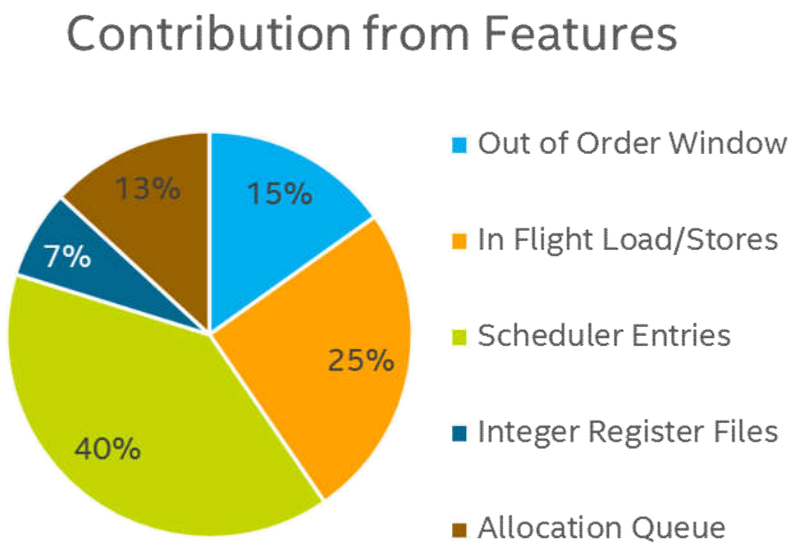
	Sandy Bridge	Haswell	Skylake
Out of Order Window	168	192	224
In-flight Loads	64	72	72
In-flight Stores	36	42	56
Scheduler Entries	54	60	97
Integer Register File	160	168	180
FP Register File	144	168	168
Allocation Queue	28/thread	56	64/thread



Extracting more parallelism each generation, ~10% IPC improvement

Cycle Per Packet Improvements

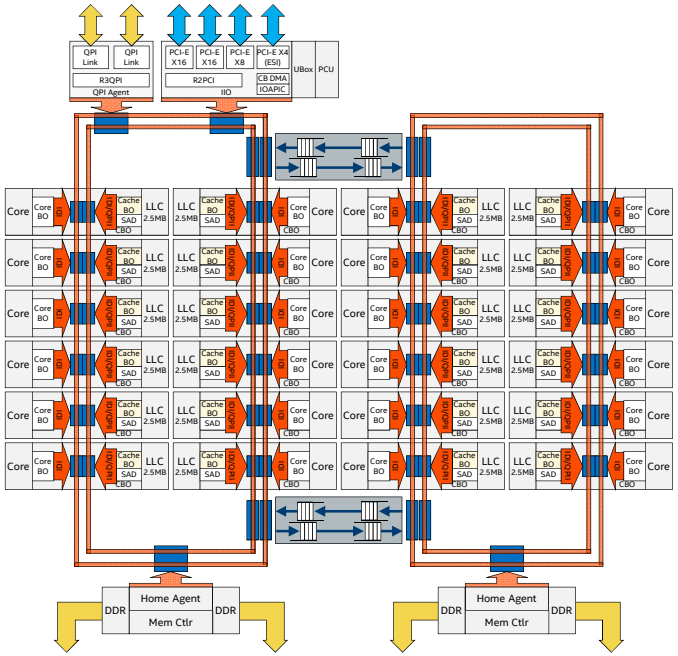




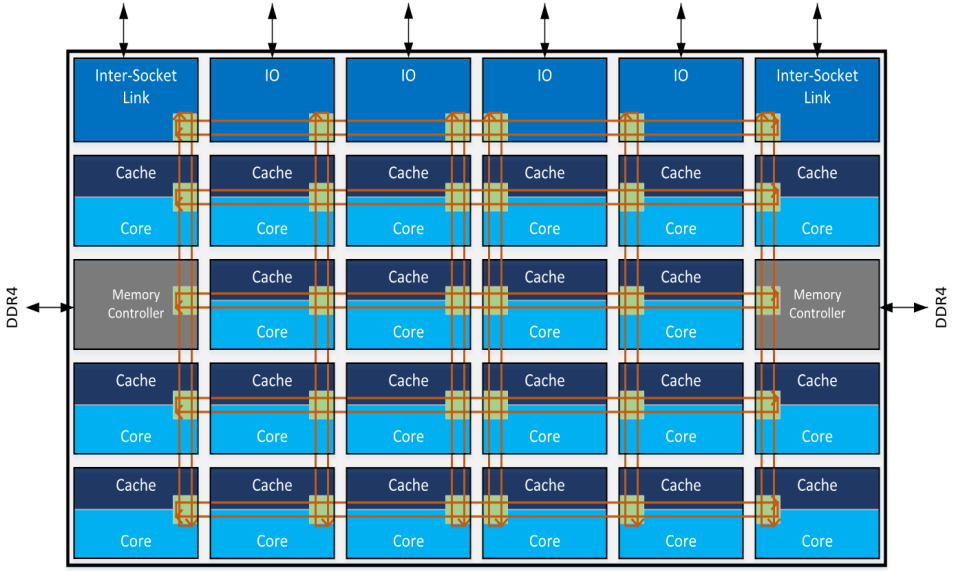
System configuration is the same as the one used in DPDK layer 3 forwarding test covered in this presentation

Skylake-SP Scalable Coherent Fabric Overview

Xeon E7 v4 24-core die

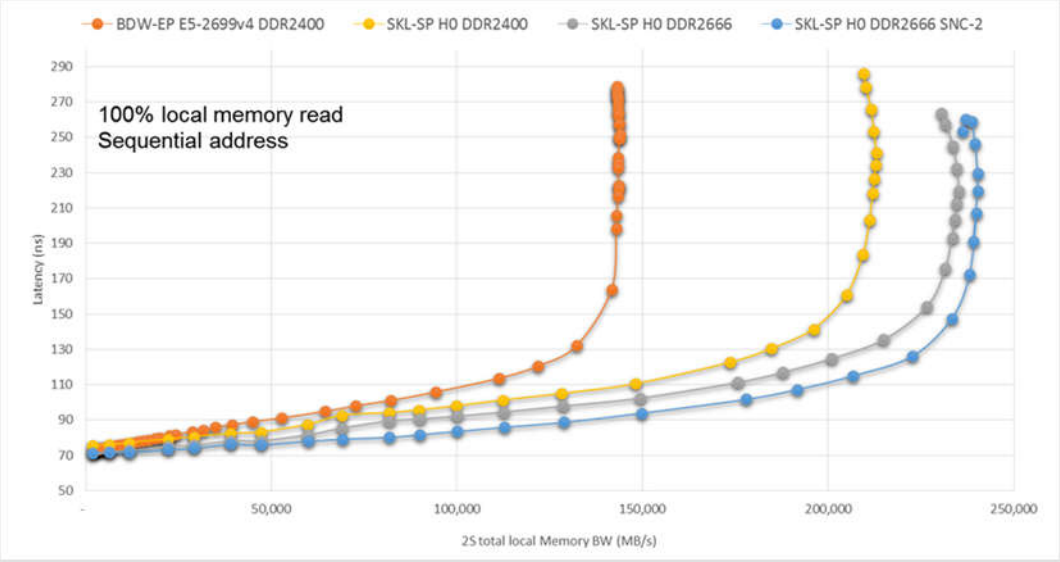


Skylake-SP



Mesh Improves Scalability with Higher Bandwidth and Reduced Latencies

Loaded Memory Access Latency

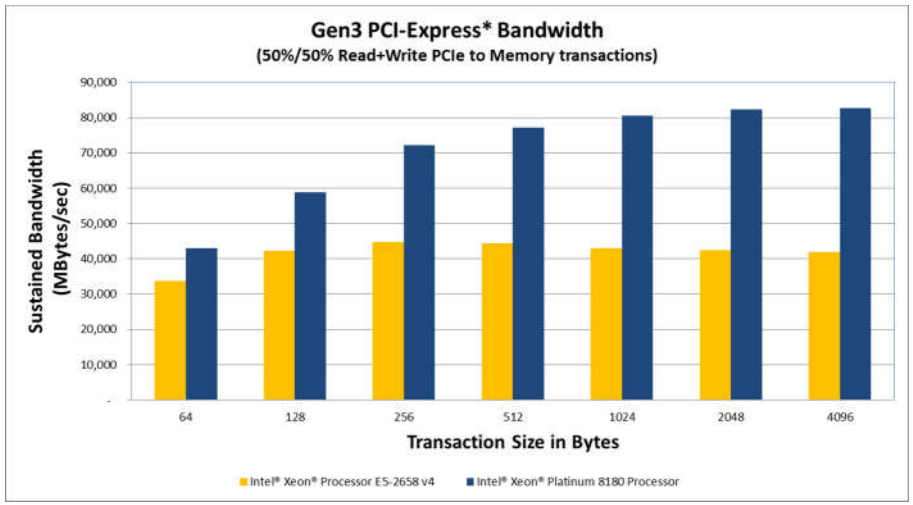


(*) Source as of May 2017: Intel internal measurements of BW/latency on platform with Skylake-SP H0 28C internal sample, Core=turbo, CLM=turbo, UPI=10.4, SNC1, 6x32GB DDR4-2400/2667 per CPU, 1 DPC, and platform with E5-2699 v4, Turbo enabled, 4x32GB DDR4-2400, RHEL 7.0. Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. **For more information go to** http://www.intel.com/performance *Other names and brands may be claimed as the property of others.

Memory Load Line enables deterministic packet processing at peak levels

- Network Function Virtualization requires deterministic throughput as VMs are added
- Memory controller design and two additional memory channels yield a significant improvement in the loaded latency

PCIe Bandwidth



PCI Express platform performance increases up to 2x

- Mesh to I/O improvement, three MS2PCI mesh stops
- Additional Gen 3 x16 PCI E interface, three in total – resulting in up to 82GB/Bytes per socket
- Improvement in Data Directed I/O architecture, separation of RX and TX data

"Results have been estimated based on internal Intel analysis and are provided for informational purposes only. Any difference in system hardware or software design or configuration may affect actual performance. Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more information go to http://www.intel.com/performance/datacenter. Configurations: see next slide"

Translating Core, Memory and I/O Performance to Packet Processing

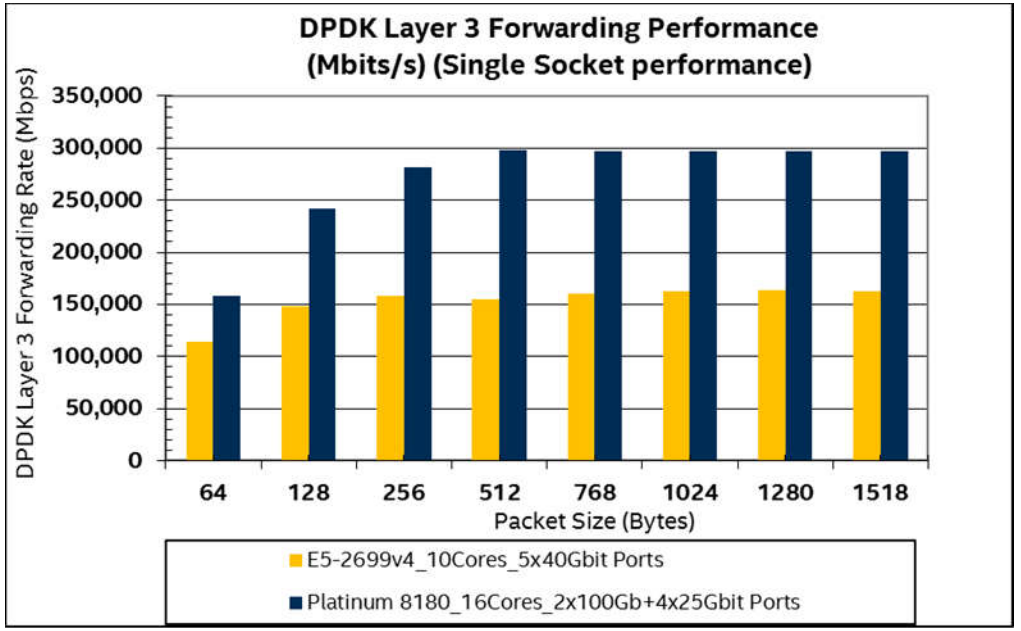
Data Plane Development Kit

Linux* Foundation Project

- More than 20 key open source projects build on DPDK libraries, including MoonGen*, mTCP*, Ostinato*, Lagopus*, Fast Data (FD.io), Open vSwitch*, OPNFV*, and OpenStack*

SKL-SP Optimizations

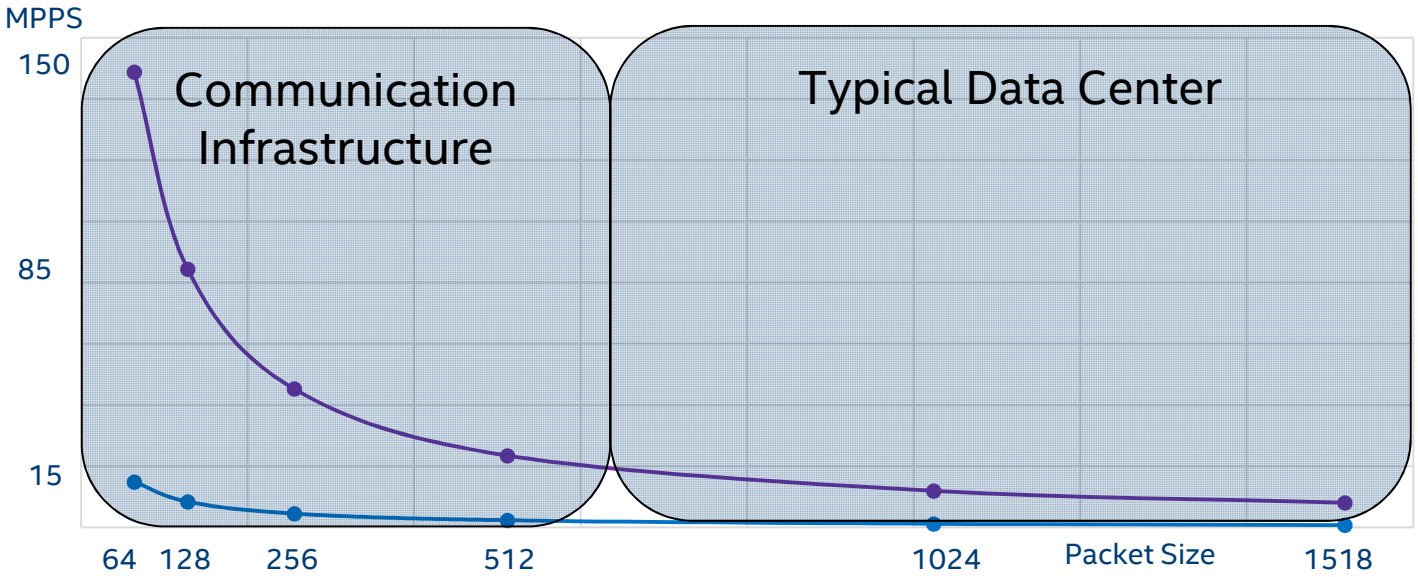
- Large MLC enables packet processing application foot print to remain close to the core



"Results have been estimated based on internal Intel analysis and are provided for informational purposes only. Any difference in system hardware or software design or configuration may affect actual performance. Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more information go to http://www.intel.com/performance/datacenter. Configurations: see next slide"

*Other names and brands may be claimed as property of others

Packet Processing Problem Statement

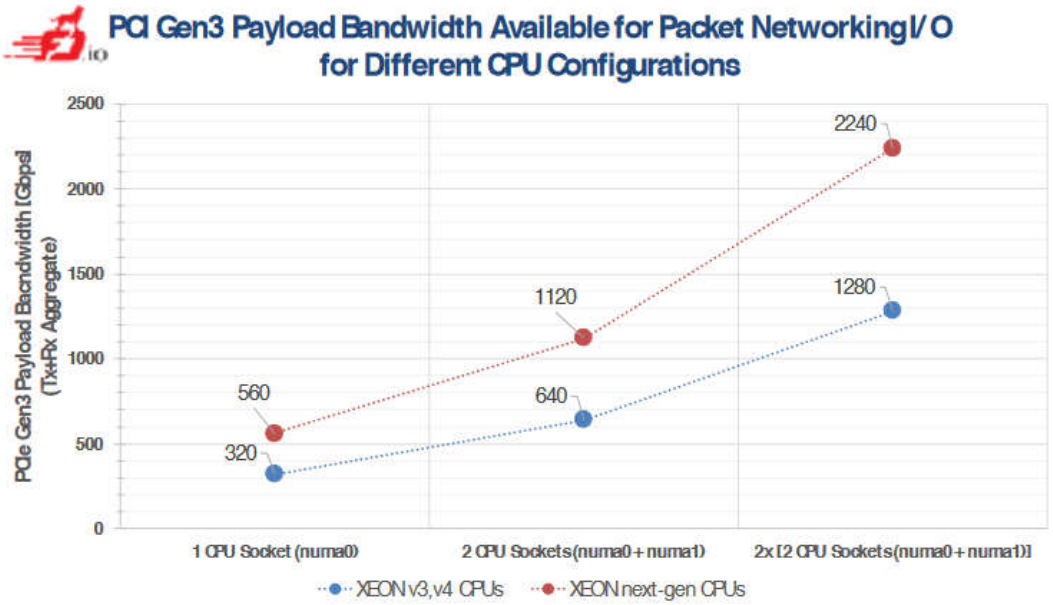


	64 Byte Packet	1024 Byte Packet
10 Gb/s	51 ns	819 ns
100 Gb/s	5 ns	82 ns

From a CPU perspective:

- Last-level-cache (L3) hit ~40 cycles
- L3 miss, memory read is ~70ns (140 cycles at 2GHz)
- Added security complexity
- Harder to address at 100Gb rates

Terabit Throughput Level with Unmodified SW



Intel® XEON® CPUs (Skylake-SP)
a. Per socket have 48 lanes of PCIe Gen3
b. 2x 280Gbps of packet I/O per socket

Intel® XEON® CPUs (E5 v3/v4)
a. Per socket have 40 lanes of PCIe Gen3
b. 2x 160Gbps of packet I/O per socket

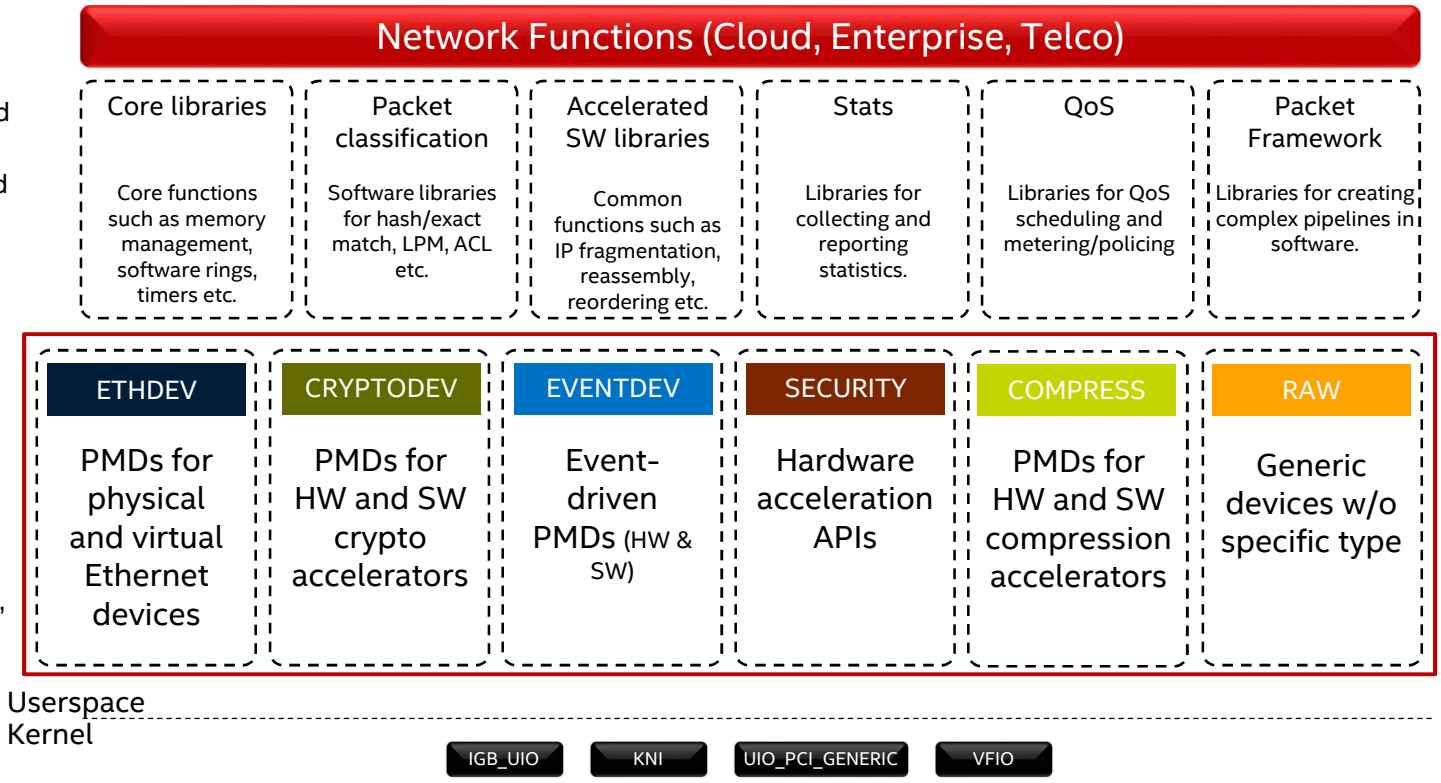
Breaking the Software Defined Network Services Barrier
1 Terabit Services on dual Intel® Xeon® Server !!! with DPDK, Fortville-25, Lewisburg

https://fd.io/2017/07/fdio-doubles-packet-throughput-performance-terabit-levels/
https://www.sinog.si/wp-content/uploads/2017/05/SINOG-VPP.pdf

Unlocking Platform Capability by DPDK

DPDK Fundamentals

- Implements run-to-completion and pipeline models
- No scheduler - all devices accessed by polling
- Supports 32-bit and 64-bit OSs, with and without NUMA
- Scales from Intel® Atom® to Intel® Xeon® processors
- Number of cores and processors is not limited
- Optimal packet allocation across DRAM channels
- Use of 2M & 1G hugepages and cache aligned structures
- Uses bulk concepts - processing 'n' packets simultaneously
- Open source and BSD licensed

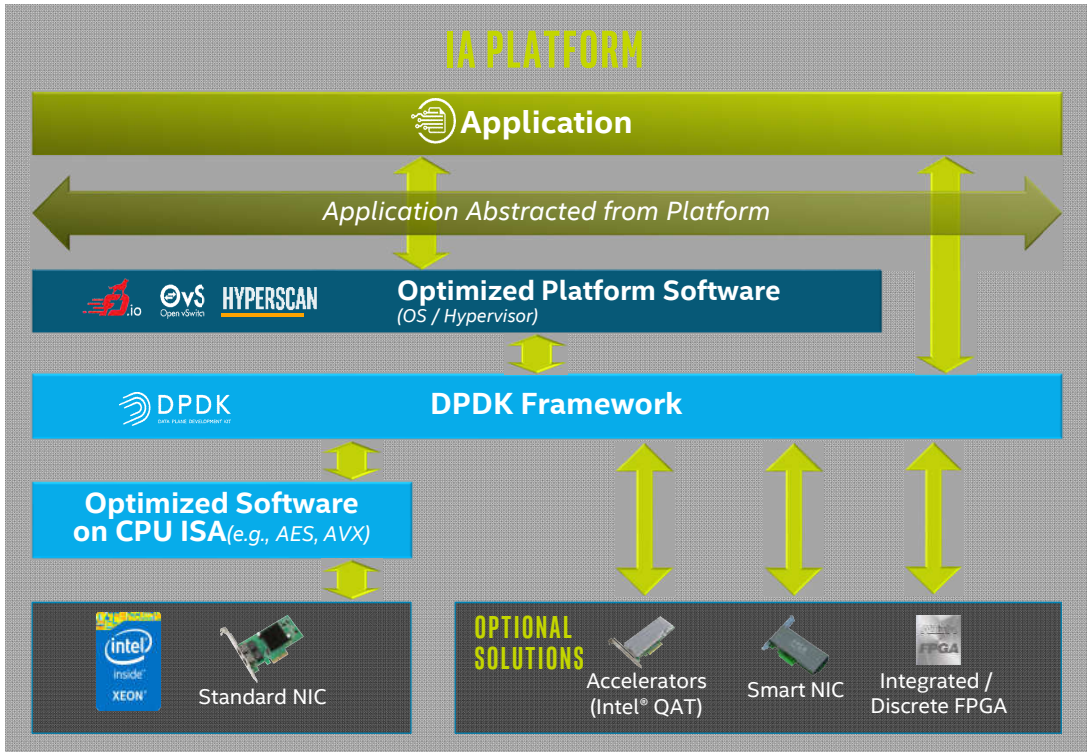


Bridging Various Accelerators

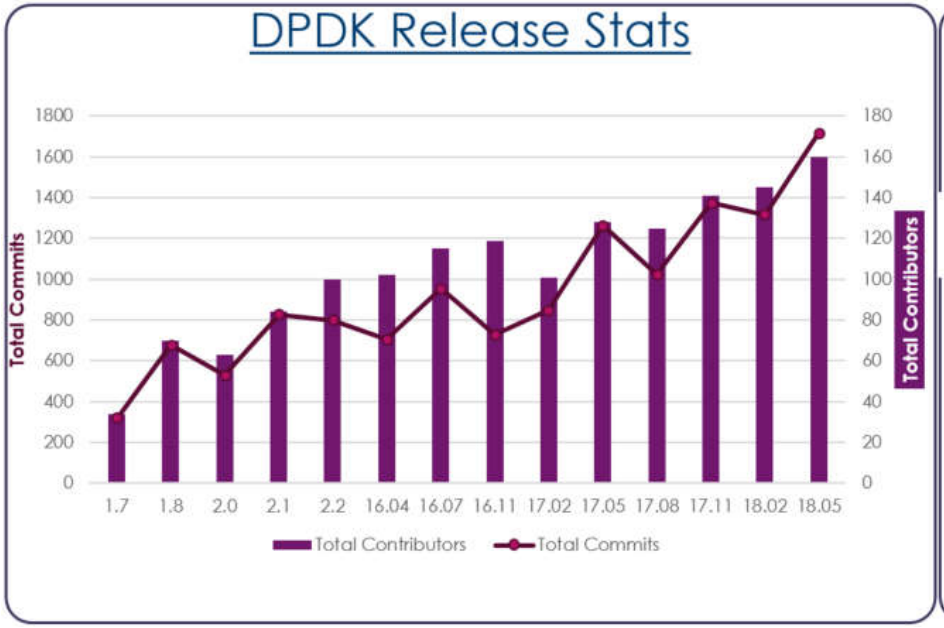
seamless interface to accelerators

DPDK Framework

- Generic APIs
- Application is abstracted from the underlying SW and HW with DPDK
- Preserve Platform and Application software investment
- Optimized platform software ingredients (e.g. vSwitch) to take advantage of HW and SW ingredients
- Flexible and outstanding performing data plane

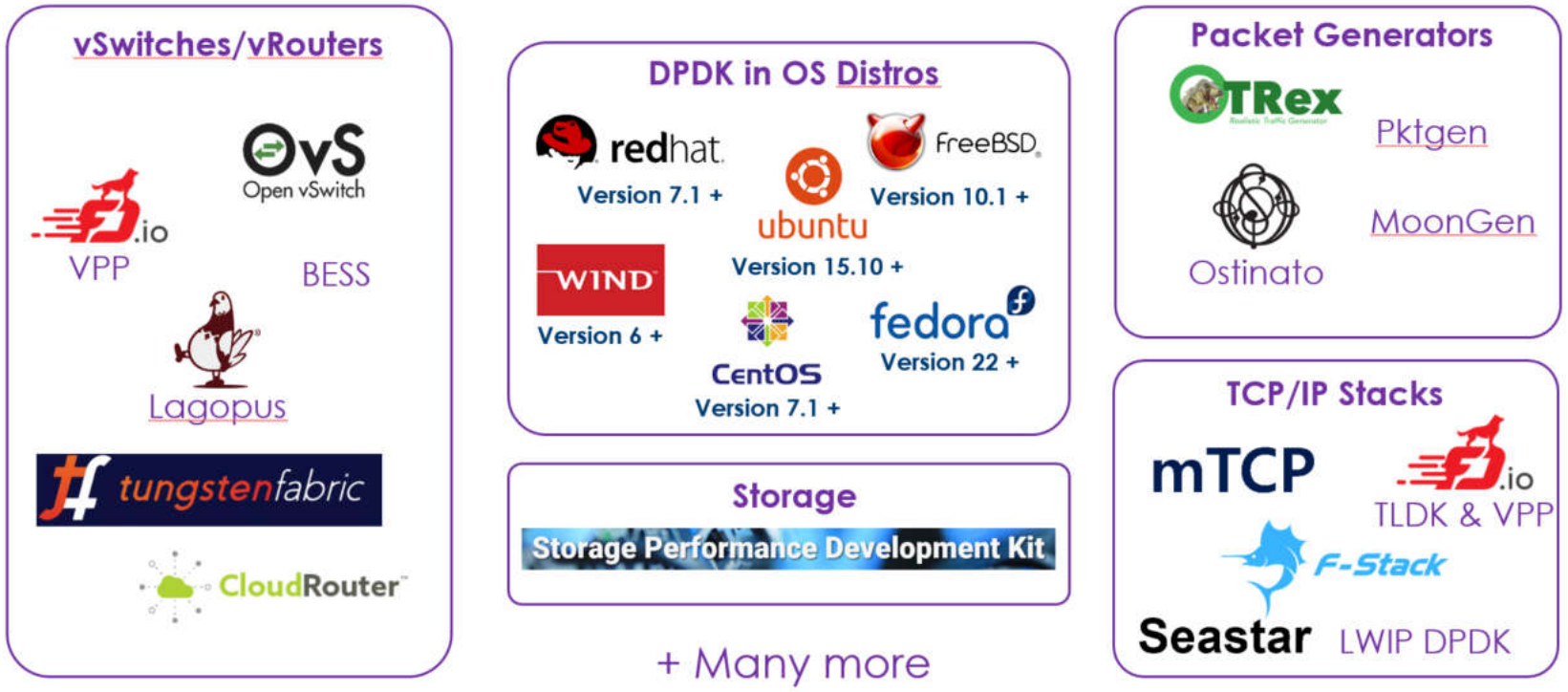


Community Ecosystem



A fully open source software project with a strong development community

Boosts Open Source Projects



Enriches Research & Innovation

MICA [NSDI '14]

Software RAN [CCTS '15]

mTCP [NSDI '14]

BlindBox [SIGCOMM '15]

Trumpet [SIGCOMM '16]

IX [OSDI'14]

ScaleBricks [SIGCOMM '15]

PISCES [SIGCOMM '16]

FTMB [SIGCOMM '15]

MoonGen [IMC '15]

ESWITCH [SIGCOMM '16]

OpenNetVM [HotMIddlebox '16]

ClickNP [SIGCOMM '16]

NetBricks [OSDI '16]

SwitchKV [NSDI '16]

NFVnice [SIGCOMM '17]

SoftFlow [ATC '16]

Flowtune [NSDI '17]

APUNet [NSDI '17]

Decibel [NSDI '17]

mOS [NSDI '17]

NetCache [SOSP '17]

StatelessNF [NSDI '17]

VigNAT [SIGCOMM '17]

NFP [SIGCOMM '17]

STYX [SOCC '17]

ExpressPass [SIGCOMM '17]

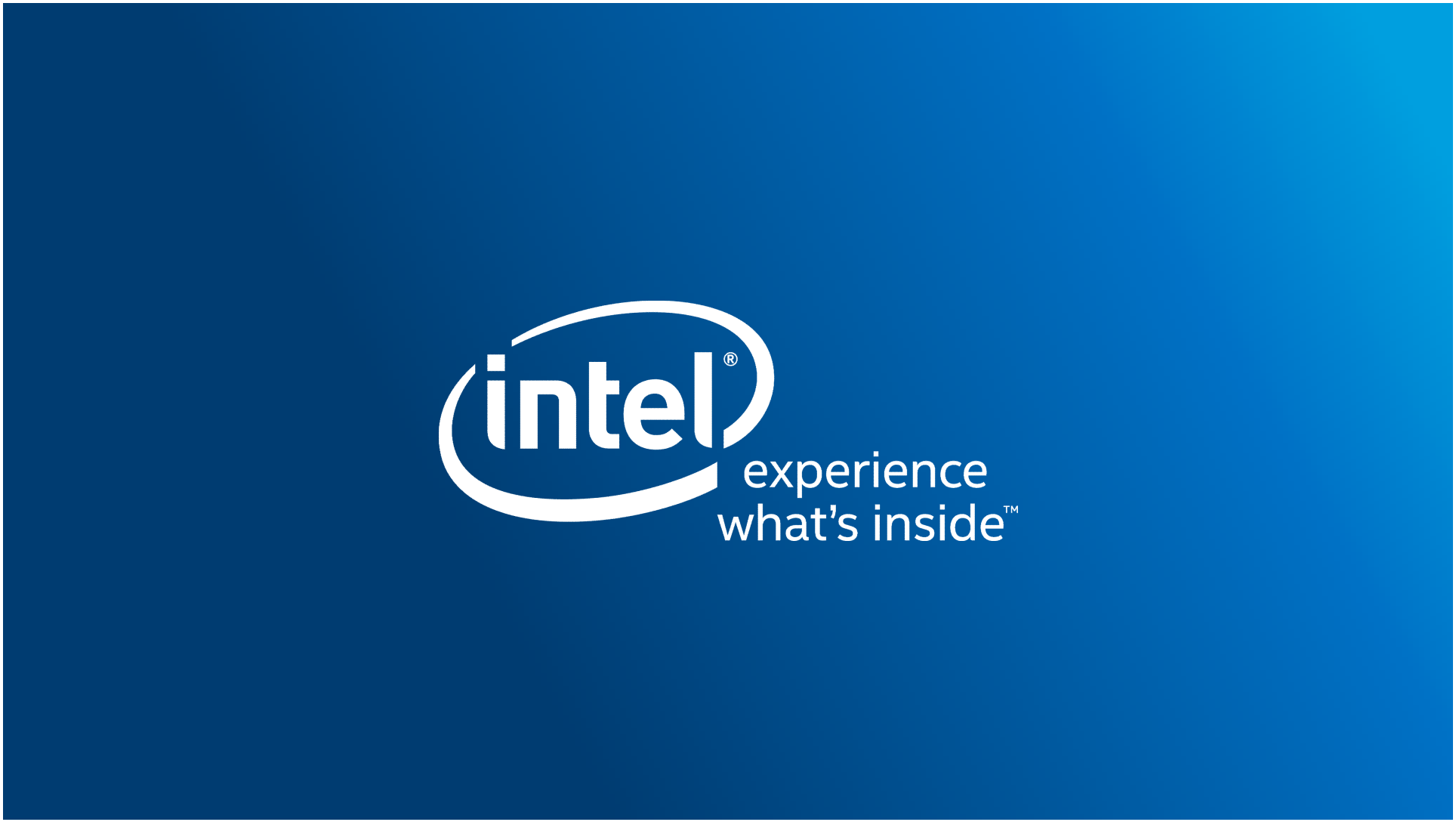
Future: Toward Cloud-Native Network Functions

- Primary Constructs
  - DevOps/Continuous delivery/Micro services/Containers

- Unique Considerations of Network Functions
  - Data plane packet processing requires an optimized architecture
  - Domain specific protocol is absent
  - Intergenerational transforming & compatibility

Summary

- Powerful Multi-Core Scalable Architecture Processor
- Unlock Packet Processing Capability by DPDK
- Seamless Interface to Various Accelerators
- Fantastic Ecosystem for Innovation


intel®
experience
what's inside™