IDF13

Extending Open Networking Platform (ONP) for the Next Generation Server Architectures

Gary Lee – Product Marketing Engineer, Intel Corporation

CLDS006

intel Look Inside.

Agenda

- Data Center Trends
- New Microserver Solutions
- Intel® Ethernet Switch Silicon Architecture
- Rack Scale Architecture
- Software-defined Infrastructure
- Summary

Key Data Center Trends

- **Dense computing resources**
  - Microservers
    - Improved efficiency for certain workloads
  - Rack scale architecture
    - Efficient data center building blocks
    - Dis-aggregation of resources
- **Software-defined infrastructure**
  - Open software for data center networks
    - Reduces CapEx and OpEx
  - Network virtualization and tunneling
    - Enables multi-tenant environments
  - Network function virtualization
    - Virtualized network appliances

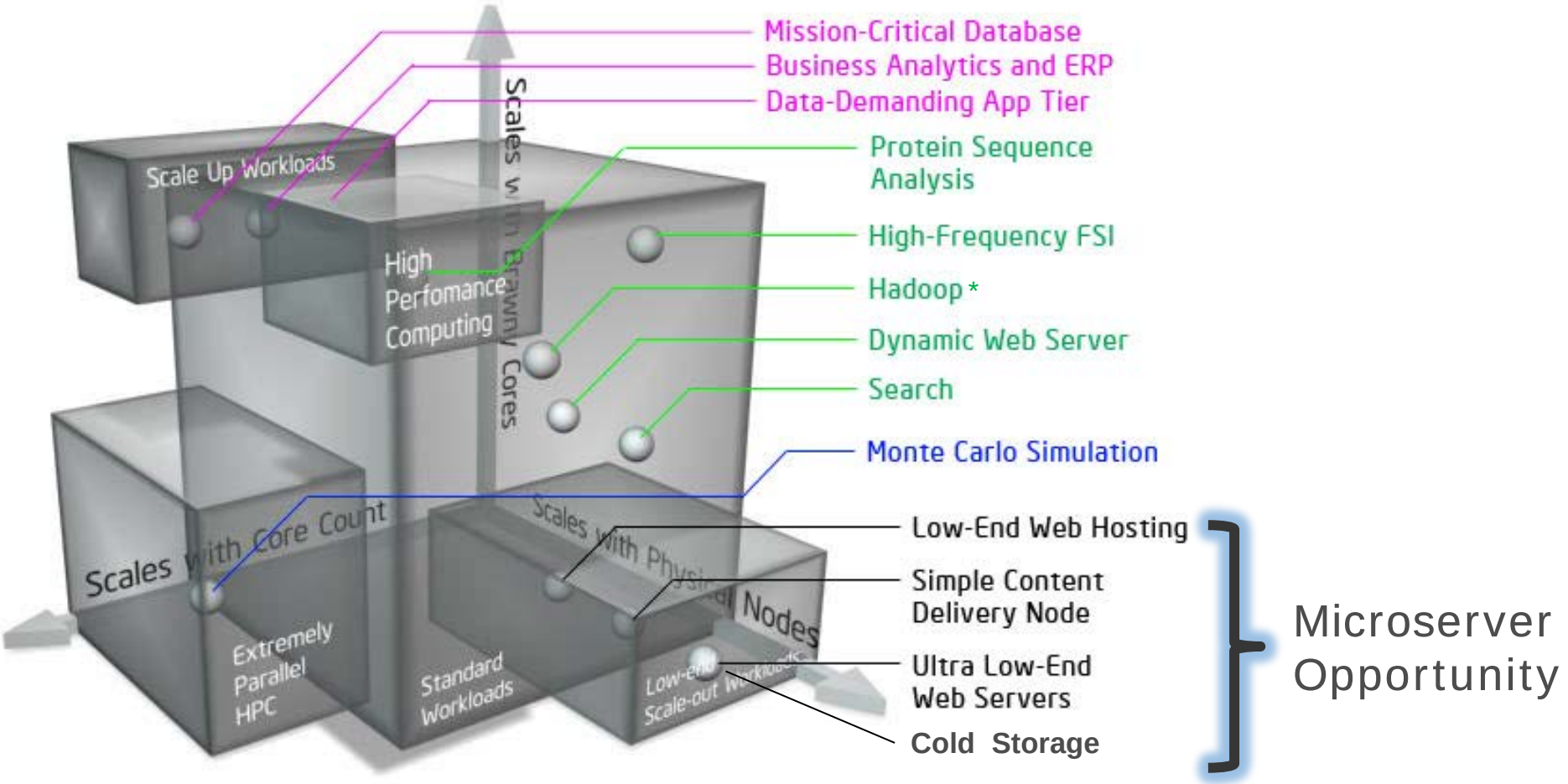
Movement to Microservers

- **Drive Perf / Watt / $ Gains**
  For unique, highly parallel workloads
- **Reduce acquisition cost**
  Per node performance less important
- **"Right size" processing**
  For lightweight, I/O bound apps
- **Maximize node density**
  Per node performance less important

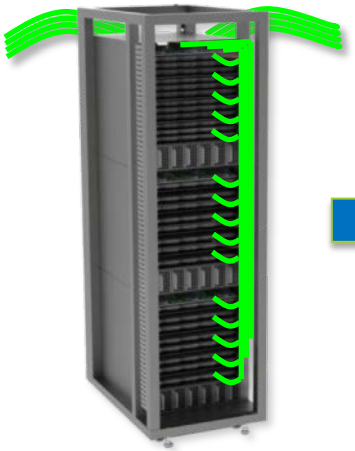
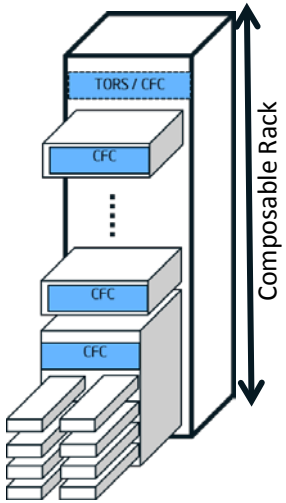
Microservers hit unique density, performance, rack & cost design targets

IDF13

Microserver Workload Opportunities

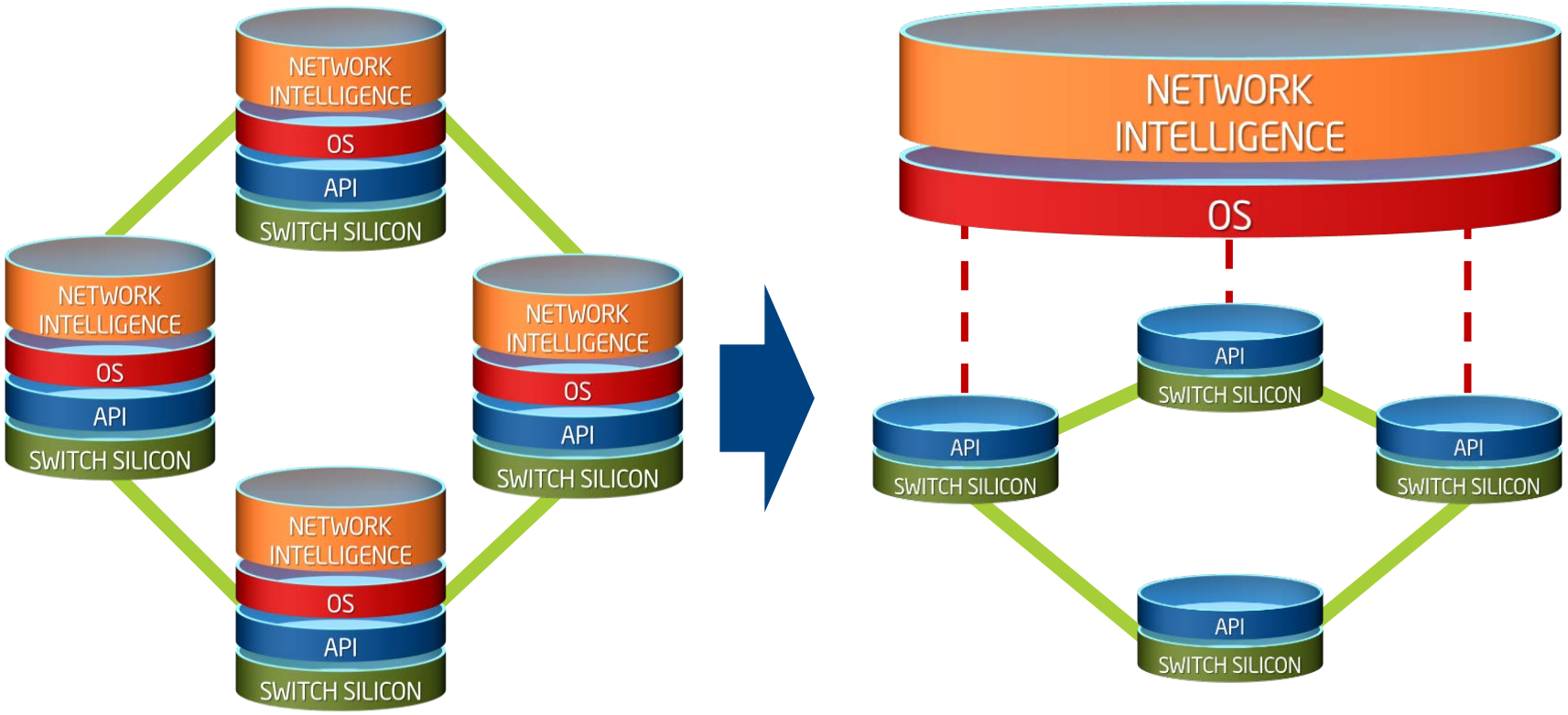


Rack Scale Architecture (RSA) Evolution

	Current: Physical Aggregation	Near Term: Fabric Integration	Longer term: Subsystem Aggregation
			
Rack level Challenges	• Power/thermal inefficiencies • Limited density	• Fixed configuration • High cable count • Limited scalability	• No service based configurability • Limited scalability
Intel RSA Solution	• Shared power • Shared cooling	• Distributed Switching • Flexible topologies • Software Defined Networking	• Pooled storage/boot • Pooled memory • Software Defined Server
End-user Benefit	• Lower TCO	• Lower TCO • Lower CPU refresh costs • Higher compute density	• Resources match workload • Service scalability

Software-defined Networking (SDN)

- Provides a centralized view of the network
- Intelligence moves from networking devices to the controller
- Reduced capital and operating expenses
- Unified control of multi-vendor network equipment
- Accelerates innovation

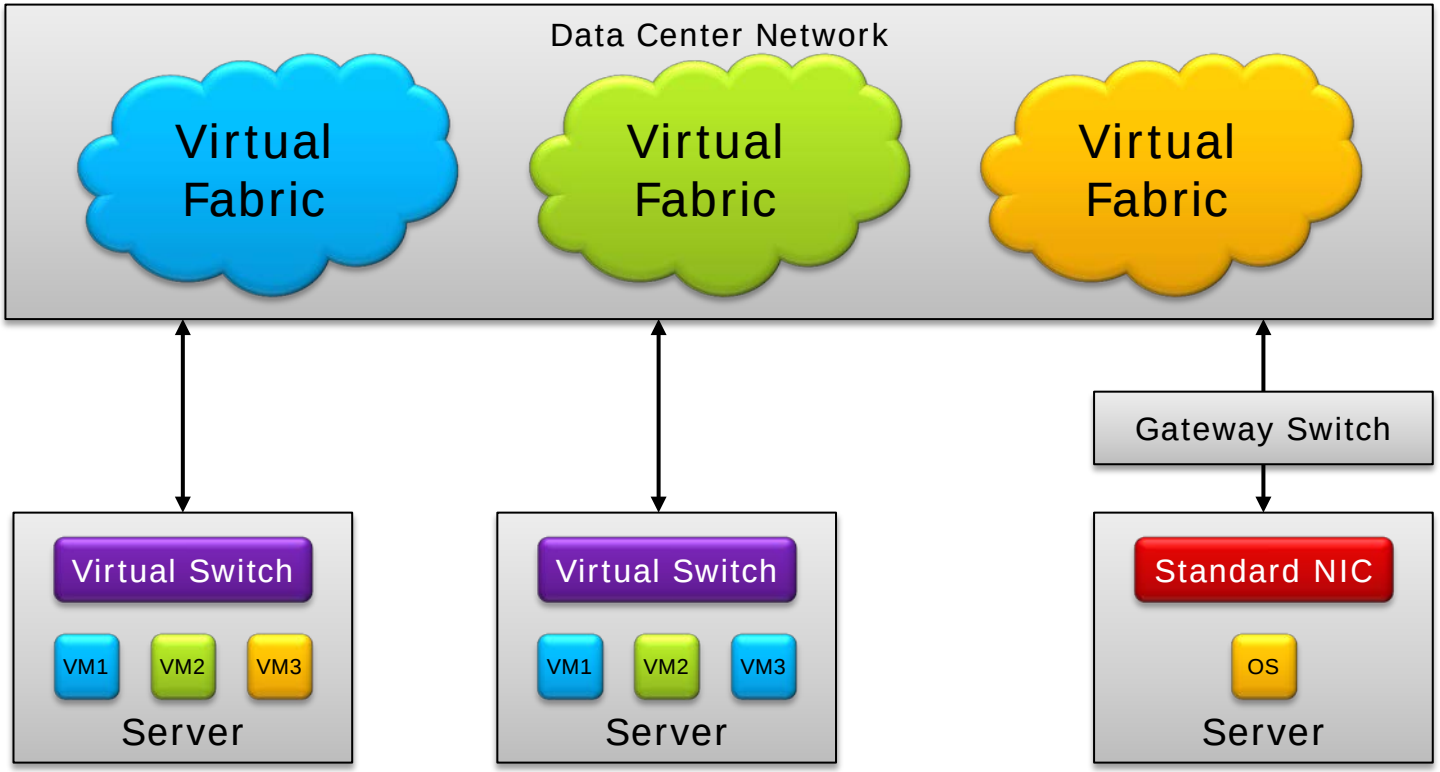


TRADITIONAL NETWORK

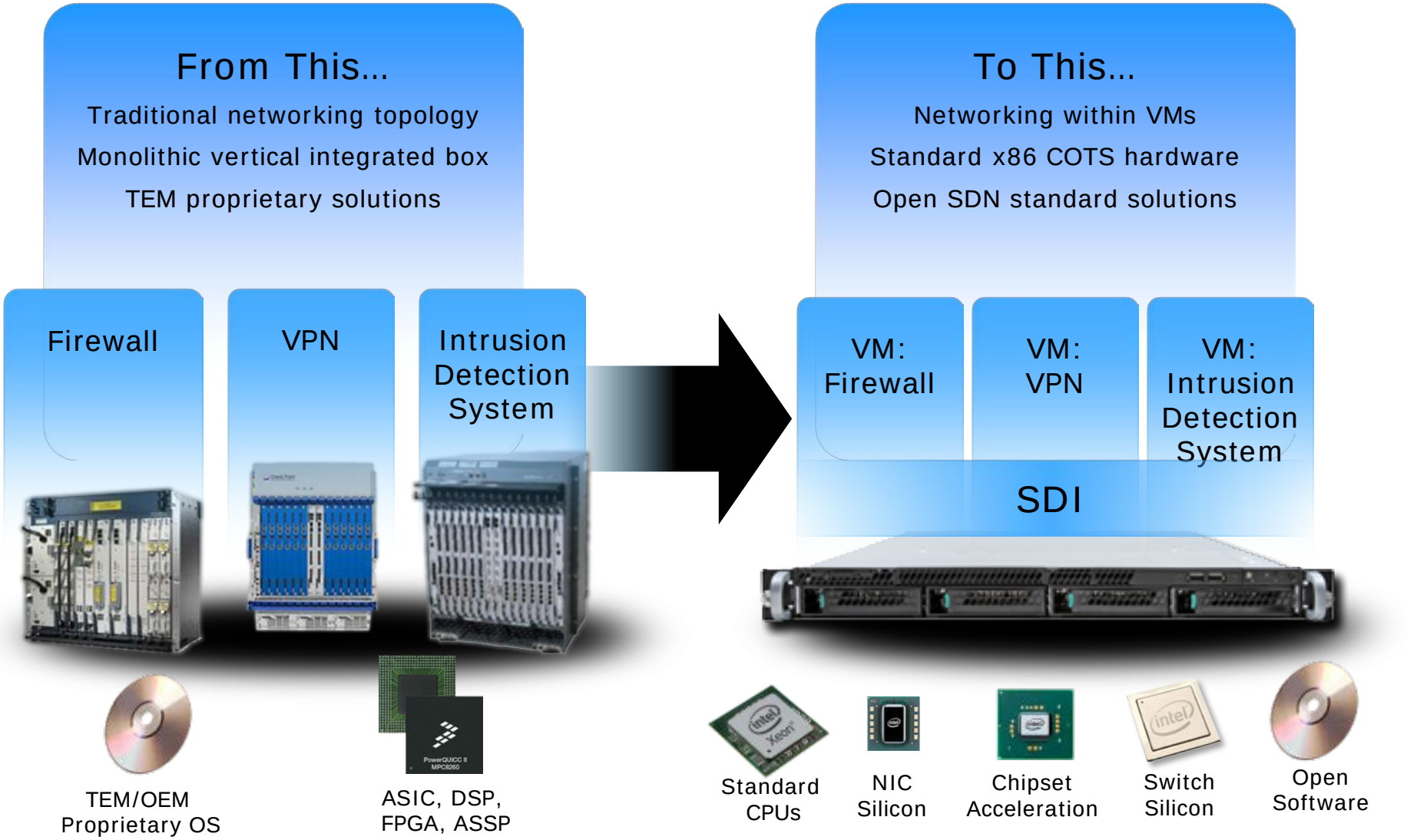
SDN NETWORK

Network Virtualization and Tunneling

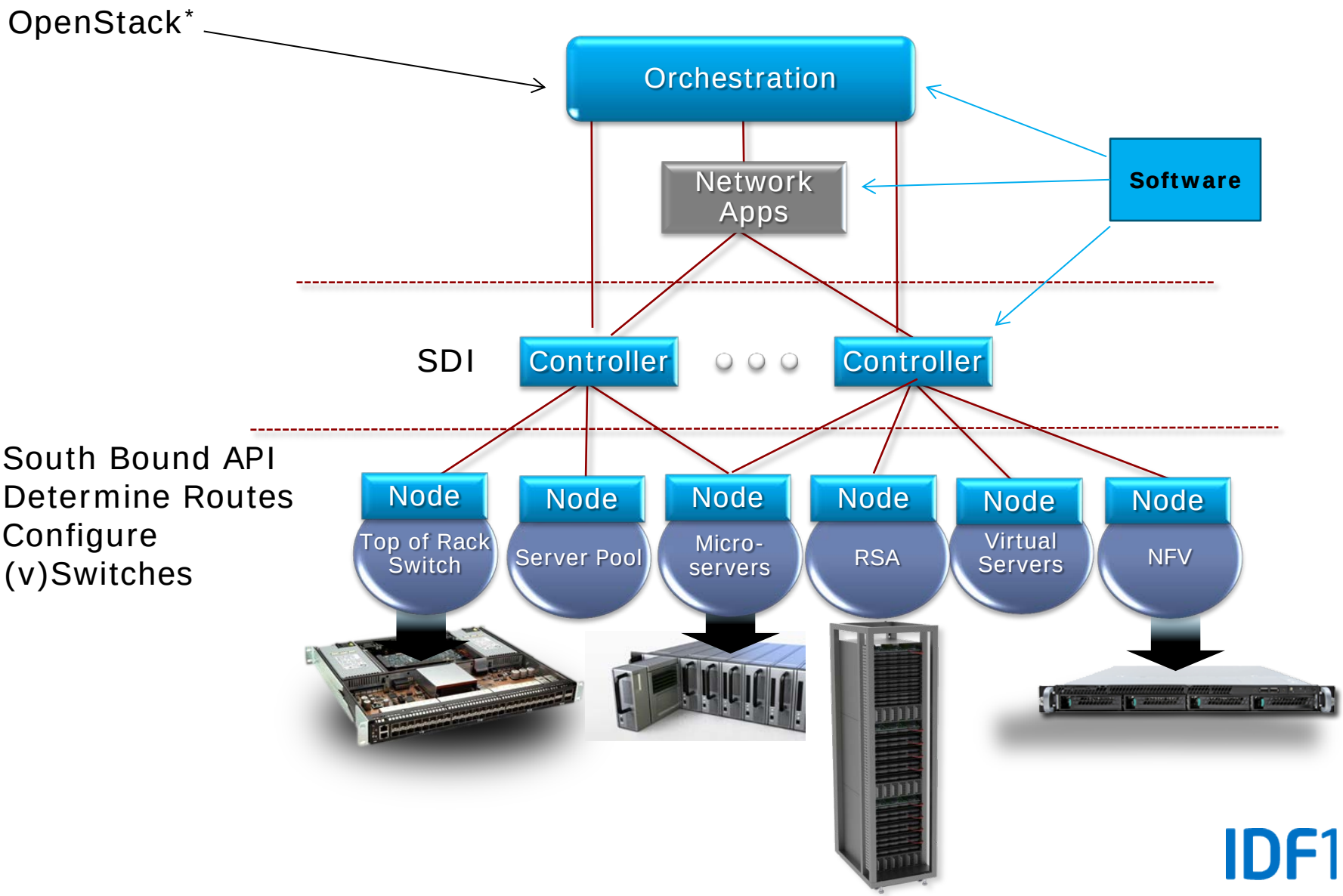
- Network Overlays provide a private virtualized data center environment for multiple tenants
- Allows flexible deployment of data center resources
- Industry standards: Virtual Extensible LAN (VXLAN), Network Virtualization using Generic Routing Encapsulation (NVGRE)



Network Function Virtualization (NFV)



How SDI Encompasses Multiple Resources



Agenda

- Data Center Trends
- New Microserver Solutions
- Intel® Ethernet Switch Silicon Architecture
- Rack Scale Architecture
- Software-defined Infrastructure
- Summary

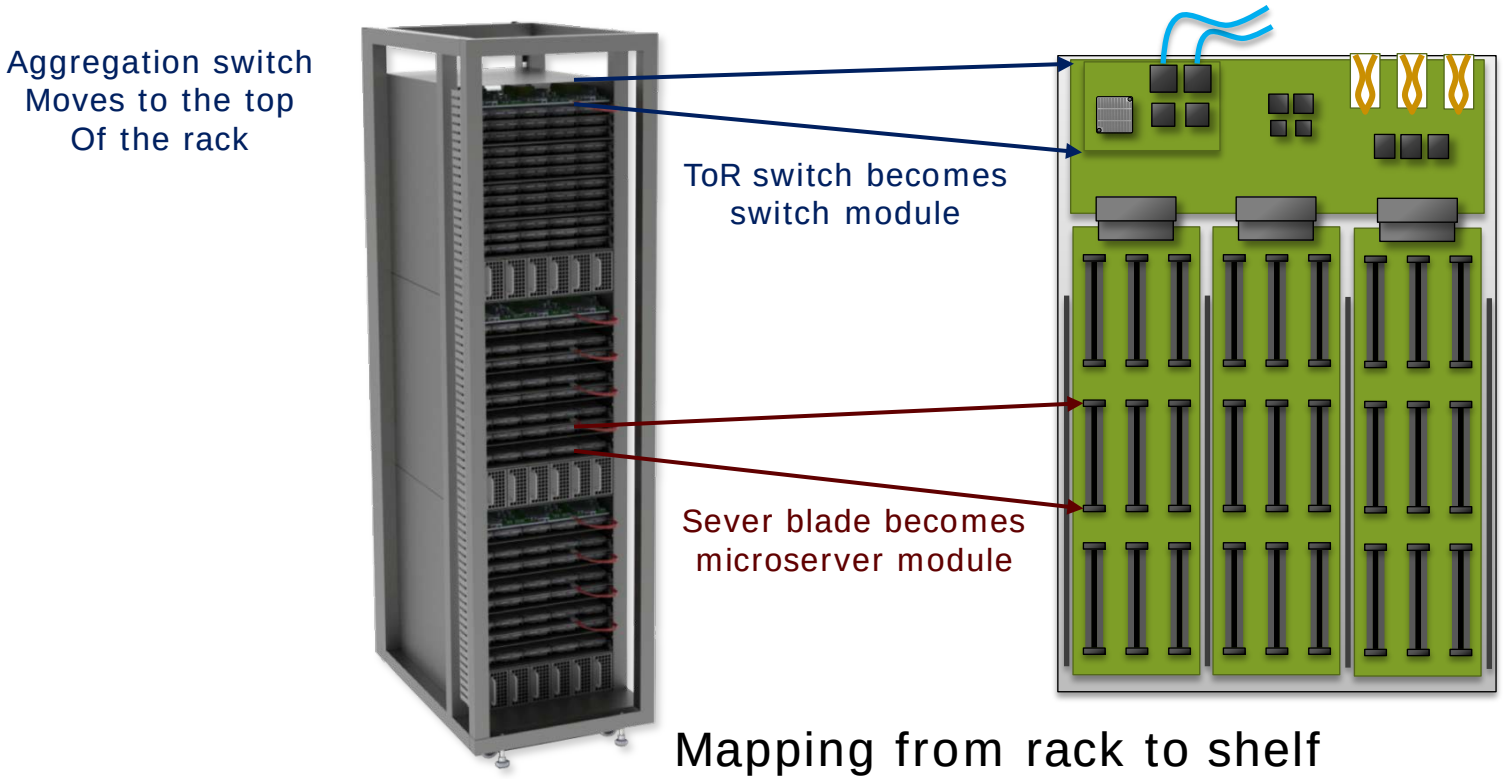
Microservers Provide High Compute Density

- Better performance per watt per dollar for unique, highly parallel workloads compared to rack servers or blade servers
  - Reduced acquisition cost
  - Right size processing for lightweight, IO bound applications
- Sharing of fans, power supplies and interconnect to achieve higher efficiency and density
  - Density: > 4X the servers per rack
  - Efficiency: Fewer fans, power supplies
  - Flexibility: Full range of 1 socket processors

Need a high port count Switch solutions on the shelf

New Switching Solutions for Microservers

- Up to 48 microservers per 3U shelf
- Up to 12 3U shelves per rack
  - 576 microservers
- Largest top of rack (ToR) switches have less than 100 ports
- Need a 2-tier solution
  - One switch per shelf with high-bandwidth uplinks
  - High bandwidth ToR switch



Mapping from rack to shelf

Microserver Switch Silicon Requirements

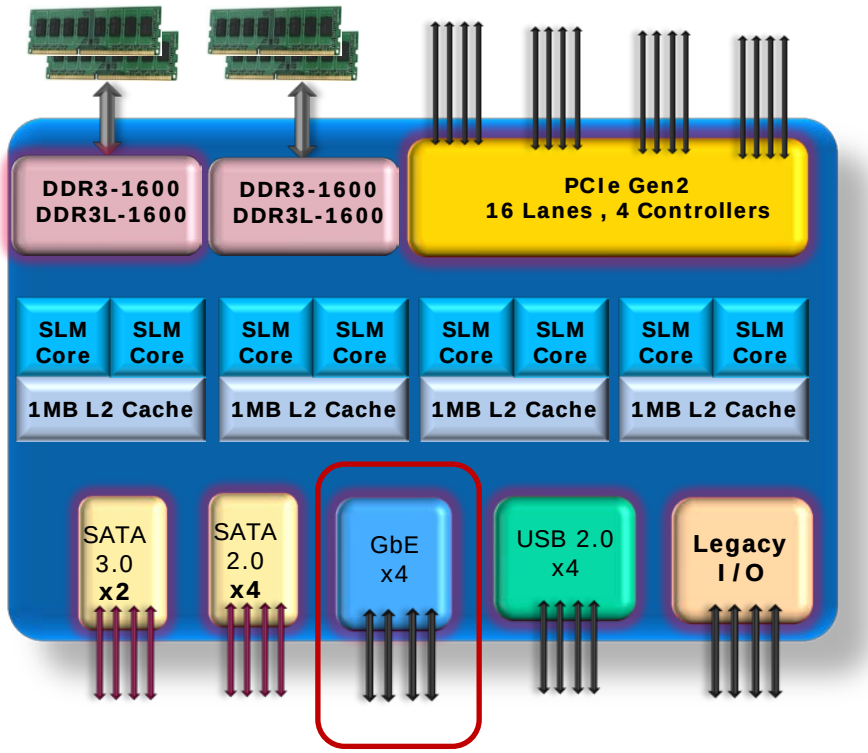
- **Bandwidth: 2.5GbE links to CPUs**
  - Allows improved performance over 1GbE solutions
- **Density: Up to 64 2.5G ports per switch**
  - Reduces the number of switch chips per system
- **Uplinks: 10G SFP+ or 40G QSFP+**
  - Uplink bandwidth must match total CPU module bandwidth
- **Software: Same as ToR switch software stack**

Microservers Need Proven System on a Chip (SoC) Building Blocks

- Need full featured Ethernet controller block
  - Virtualization support: VMDq
  - Time stamping
  - Adaptive power management: EEE
  - Broad OS support and validation
  - Support for software-defined networking
  - Advanced manageability features
- Many off-the-shelf blocks don't provide this
  - Industry proven controller block are a better choice
  - Software and OS support are key factors

Intel® Atom™ C2000 Processor Family Codename Avoton

CPU	2-8 core Intel® Atom™ Processors (22nm)
L2 cache	1MB shared per 2 cores
L1 cache	32K data, 24k instruction per core
Frequency	Up to 2.4GHz nominal (2.6GHz Turbo)
Addressing	36bit PA / 48bit VA
Memory	DDR3 up to 1600, DDR3L up to 1600, 2 channels, 2DPC, 2 ranks/DIMM, 64GB†, demand and patrol engine
Memory Types	SODIMM, UDIMM, VLP UDIMM w/ ECC
Memory RAS	Enhanced ECC SEC-DED covers address and data paths, DDR scrambler to reduce error rate, Error injection with address/source match, HW based demand and patrol engine
PCI Express* (PCIe*) Gen 2	x16 lanes , 4 controllers
Integrated IO	**4x1 GbE or 4x 2.5GbE**, x4 USB 2.0, x4 SATA 2, x2 SATA3, Server class GbE x4 ports
Technologies	Intel® VT-x2, Core RAPL, PECI over SMBUS
Targeted TDP	~5W to <20W (SKU/configuration dependent)
Package	FCBGA, Ball Pitch: 0.7mm variable Dimension: 34mm x 28mm
Legacy IO	SPI for boot flash, SMBus, UART LPC, GPIO, 8259, IO APIC, 8254 Timer, RTC



SLM= Intel Silvermont microarchitecture
Different SKUs will have different set of features
†64GB opportunistic with 8Gb densities

IDF13

Intel® Ethernet Integrated Controller IP

External Interfaces

- Integrated Quad GbE MAC/PHY/SerDes
- 1000Base-KX / 2500Base-KX

Intel® Virtualization Technology for Connectivity

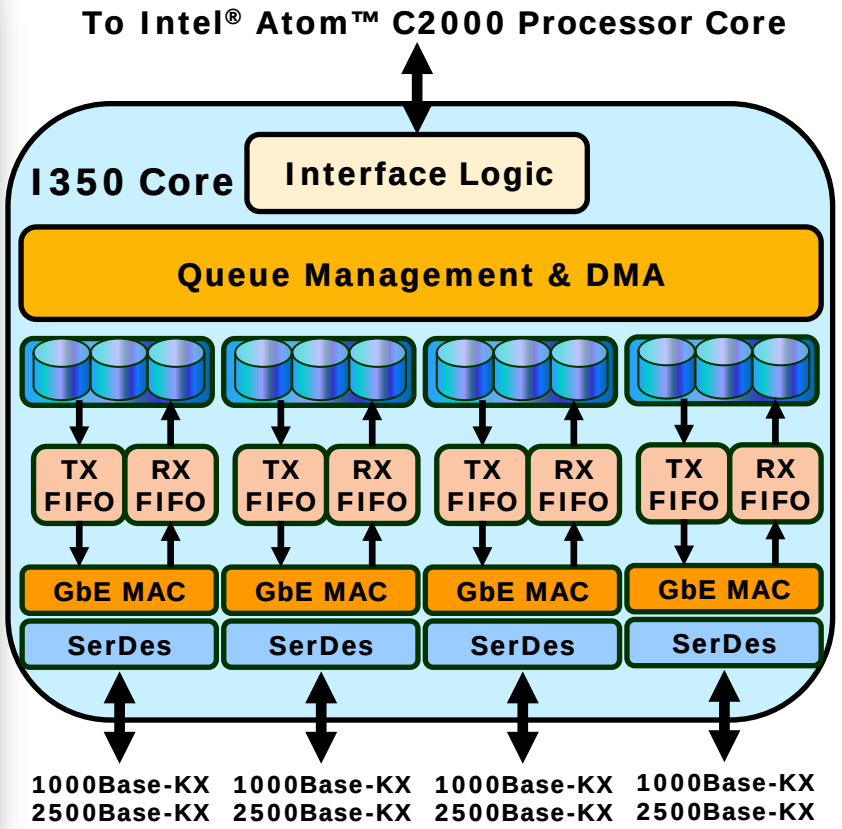
- Virtual Machine Device Queues (VMDq)
  - 8 Tx/Rx Queues (per port)
- Virtual Machine Direct Connect (VMDc)
  - 8 Virtual Functions/port with Intel® VT-d enabled

I/O Enhancements

- Max queues: 8 Rx/Tx queues/port
- Intel® I/O Acceleration Technology (stateless offloads, MSI-X, Receive Side Scaling, Low Latency Interrupts)
- IEEE 1588/802.1AS (pre-Standard) time stamping
- Protocols: TCP, UDP & SCTP

Manageability

- Power management features (PPM and EEE)
- PXE, iSCSI boot



Performance Advantages with 2.5GbE

- Reduced congestion with a high port count switch
  - FM5224 has up to 64 2.5GbE ports
  - Provides headroom for traffic bursts
- Lower latency compared to 1GbE
  - Can effect overall system performance
- 1GbE can lead to limitations on key workloads
  - Current workload requirements are less than 2.5GbE
  - MemcacheD example: Can be network limited at 1GbE on peak requests per second

Agenda

- Data Center Trends
- New Microserver Solutions
- Intel® Ethernet Switch Silicon Architecture
- Rack Scale Architecture
- Software-defined Infrastructure
- Summary

Intel® Ethernet Switch FM5224

Microserver Switch Silicon

Unmatched uServer density

- Up to 72 2.5G ports
- 8 10GbE or 2 40GbE uplinks

Rapid Array shared memory

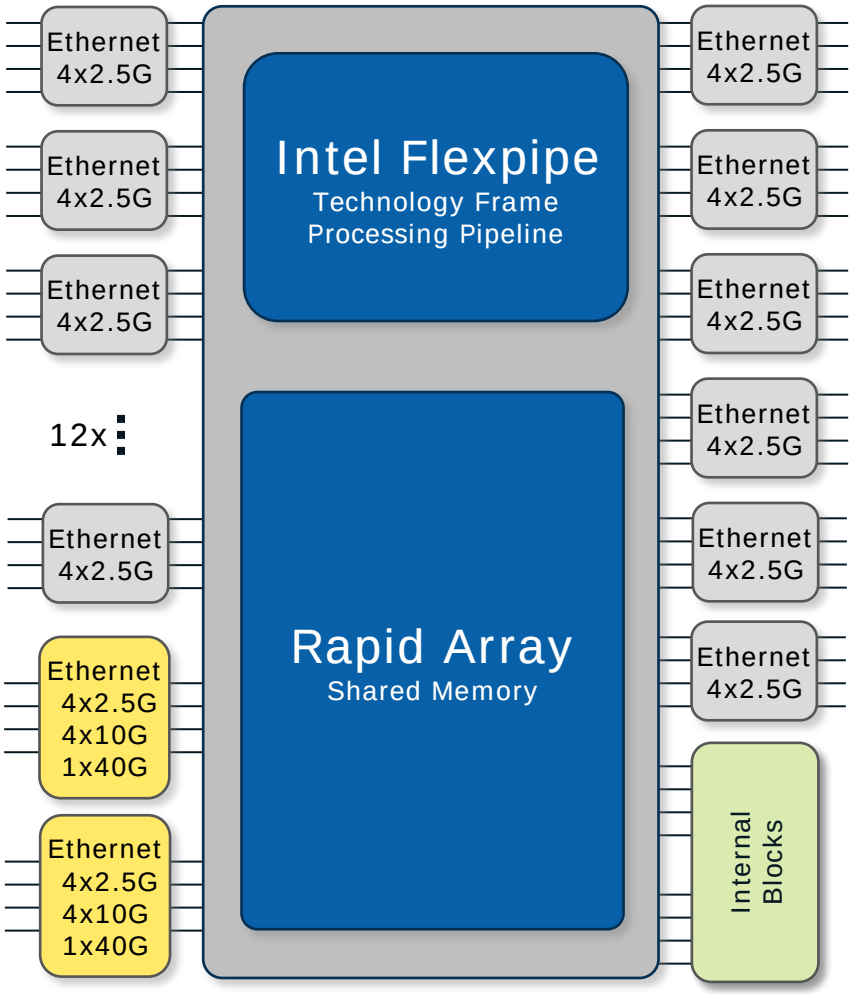
- 8MB shared memory
- 400nS cut-through latency

Intel® Flexpipe™ Technology frame processing

- Intel Flexpipe Technology frame processing
- VXLAN and NVGRE support
- Advanced load balancing
- IPv4/v6 routing
- CEE/DCB with 8 traffic classes
- Server virtualization support

Compact, flexible port logic

- Integrated SFI, KR PHY
- All ports can also operate at 10/100/1000/2500



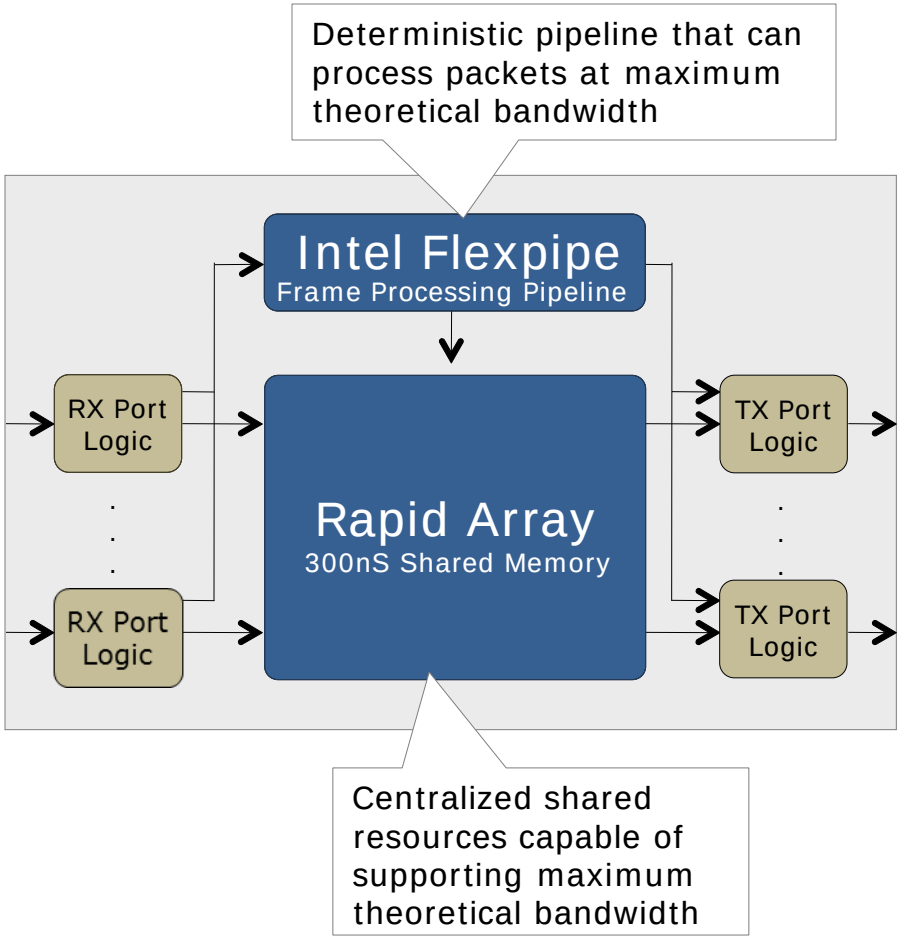
Intel® Ethernet Switch FM5224
Switch Architecture

Unmatched performance
- Full line-rate packet processing
- Industry leading L3 latency
- Fully provisioned

Rapid Array shared memory
- Low latency scheduler

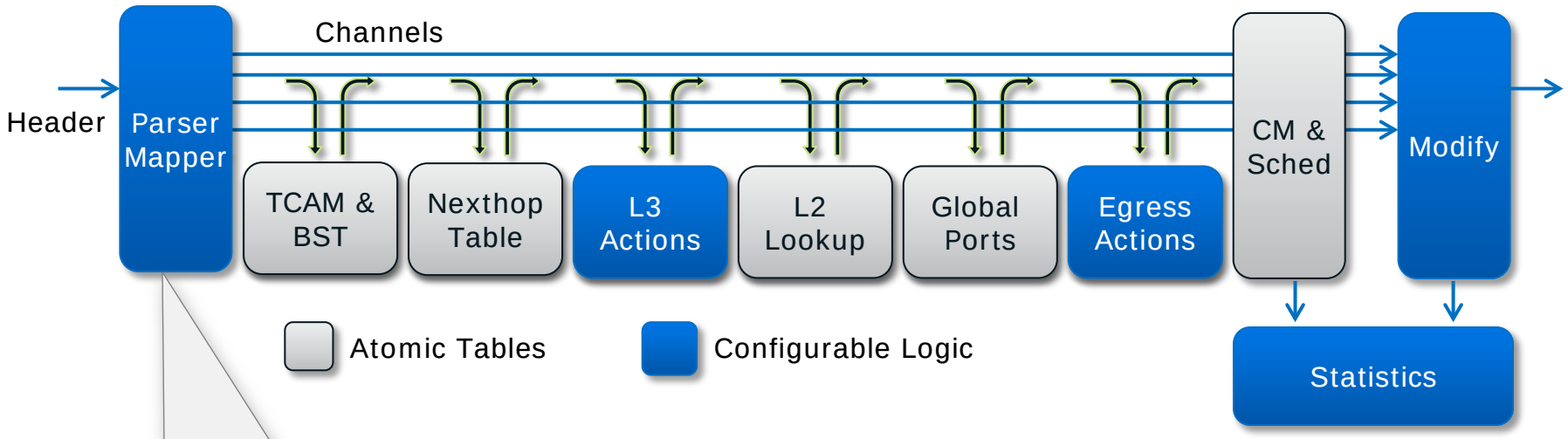
Intel® Flexpipe™ Technology frame processing
- Flexible header processing and frame forwarding
- Ideal for SDN applications
- Deterministic, programmable

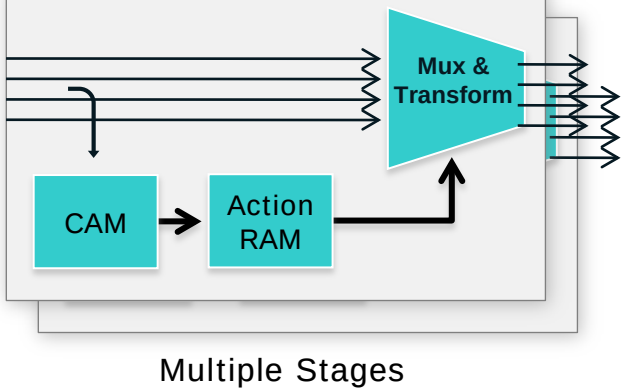
Compact, flexible port logic
- Integrated SFI, KR PHY



Output queued shared memory architecture

Intel® Flexpipe™ Technology Frame Processing Pipeline



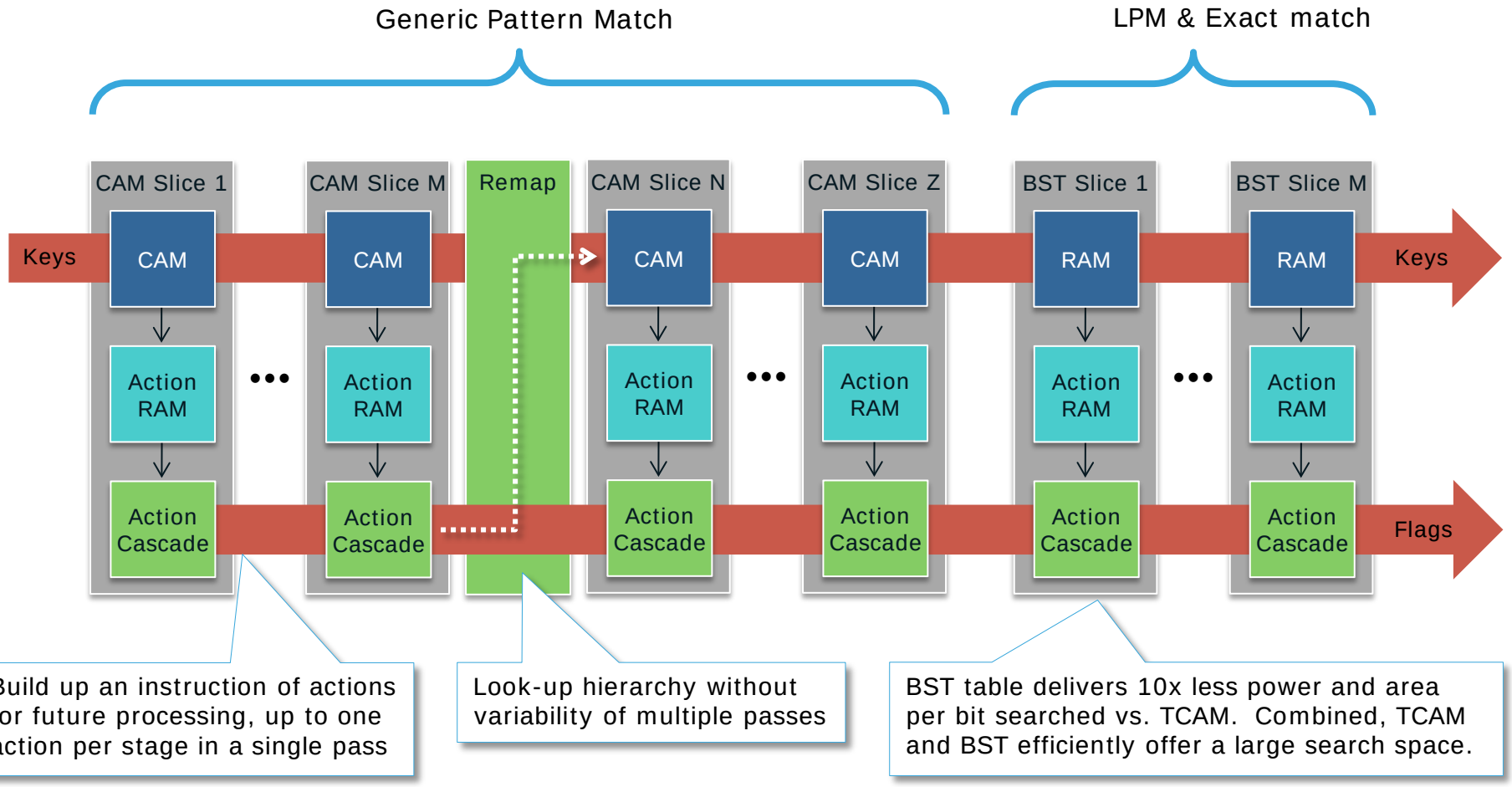


Sample Programmable Protocols

Tunneling	TRILL, MPLS, NAT
Network Overlays	VxLAN, NVGRE
Virtualization	EVB, VEPA, VEPA+, VN-Tag
Proprietary	Customer defined headers

Programmable and deterministic up to 960Mpps

The Intel® Flexpipe™ Technology TCAM / BST Architecture



Efficient filtering and forwarding using CAM and BST

IDF13

Intel® Ethernet Switch FM5224 Advanced Load Balancing Features

L2-L3 Hash-based load balancing

- Up to three 16-bit hash keys selected by ACL rules per frame
- 10 L3 or 6 L2 header fields can be used to form keys
- Symmetric hashing support

Multiple fail-over modes

- Failover to Standby
- Rolling Standby
- Redistribution to group

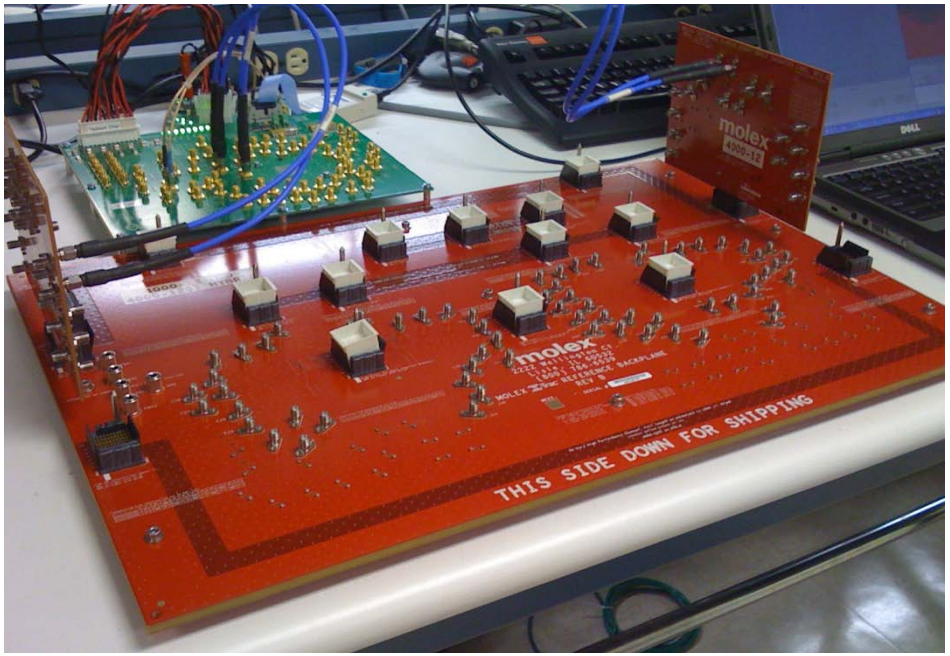
Dynamic weighted distribution

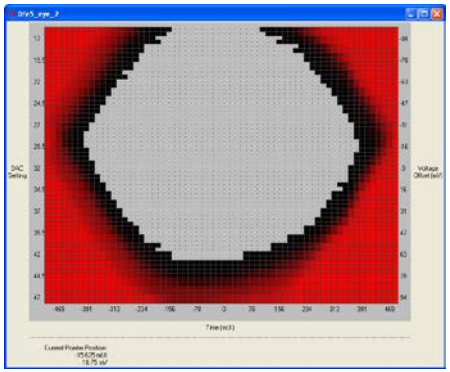
- Weights (%) assigned to each port based on 16K bins
- Supports asymmetric distribution
- Per flow statistic counters help monitor load
- Dynamic rebalancing of load

Load balance across virtual networks and virtual servers

Build-in PHY Technology

- High quality PHY technology tested across FR4 backplanes up to 10Gbps
- 2.5Gbps links verified with microserver modules, using Intel® Atom™ C2000 processor family

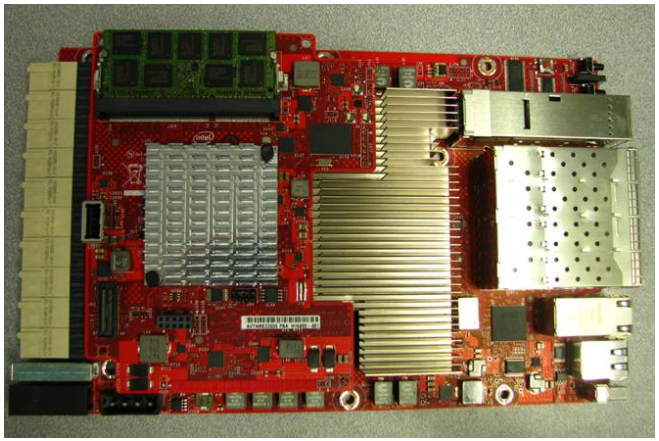




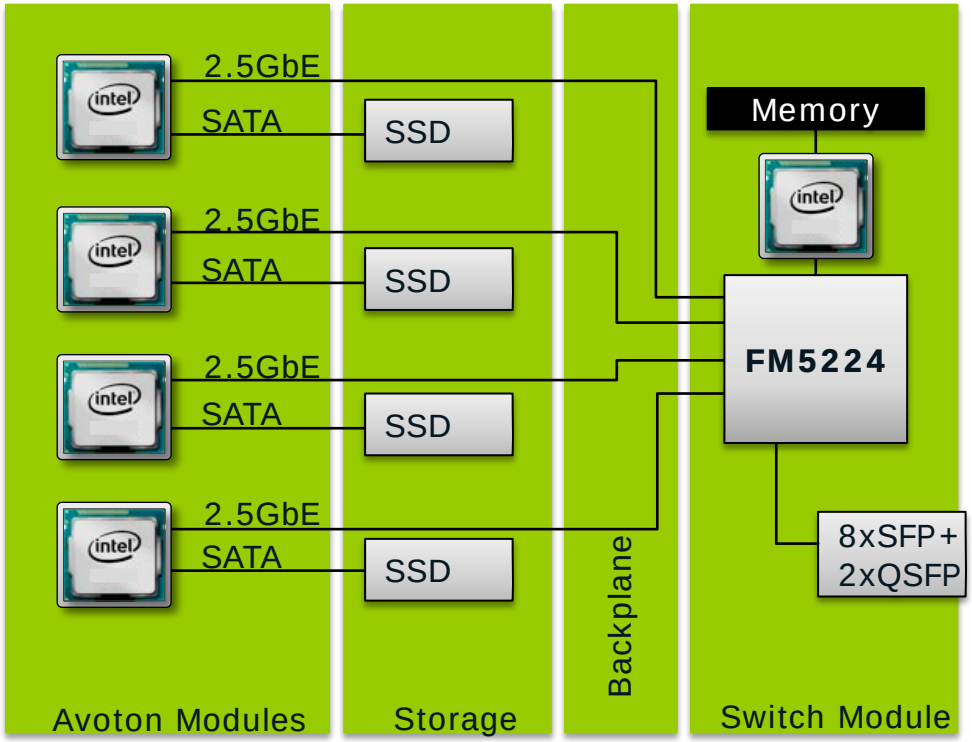
Exceptional performance and interoperability

Intel® Ethernet Switch FM5224 Provides High-Density Solutions

- **Up to 16 Avoton modules**
- **Up to 16 SSD modules**
- **Single Switch module - FM5224**
  - Avoton control plane processor
  - 8x 10GbE or 2x 40GbE uplinks
  - 64x 2.5GbE downlinks to Avotons



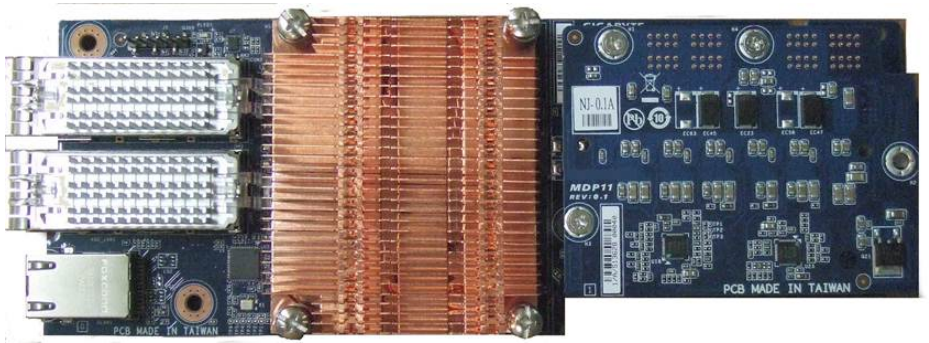
Microserver switch module SSI Form Factor



IDF13

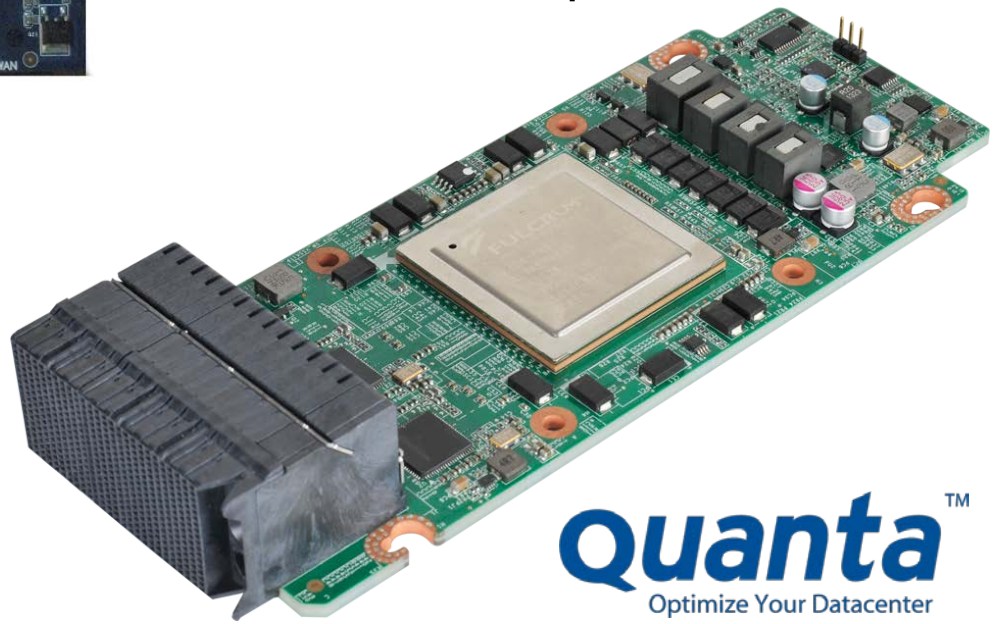
Intel® Ethernet Switch FM5224 System Deployments

NEC



- 46 2.5G links to microservers
- Two 40GbE uplink ports
- Intel® Atom™ C2000 control plane CPU

- 42 2.5G links to microservers
- Two 40GbE uplink ports
- Intel Atom C2000 control plane CPU



Quanta™
Optimize Your Datacenter

IDF13

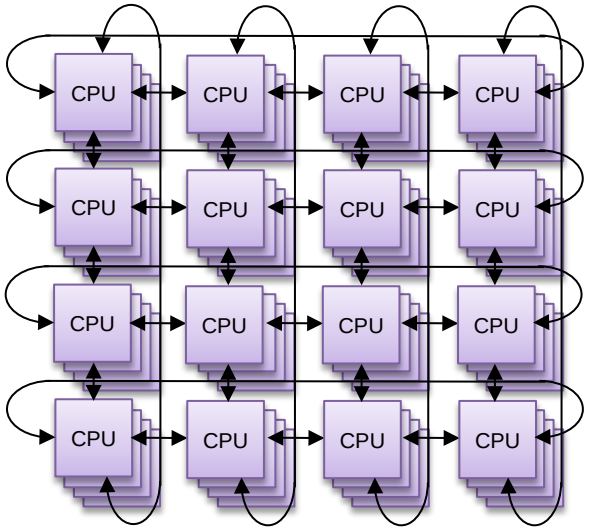
Comparison With Embedded Fabrics

Embedded solution

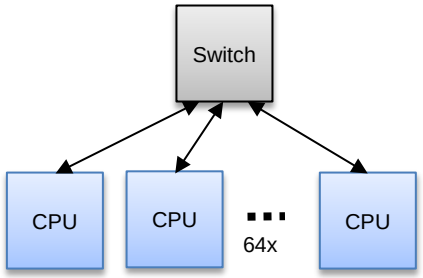
- Up to three hops between CPUs
  - Shared 1GbE bandwidth per ring
- Some solutions use additional fabric chips per CPU

Intel Solution

- Two hops between any two CPUs
  - Low latency and latency variation
- 2.5GbE links
  - Reduced congestion and lower latency
- Single switch for up to 64 CPUs
  - Can help reduce system cost
- Open network platform
  - Complete SDN solution
  - Tunneling and load balancing features



64 Microservers using Embedded Fabric



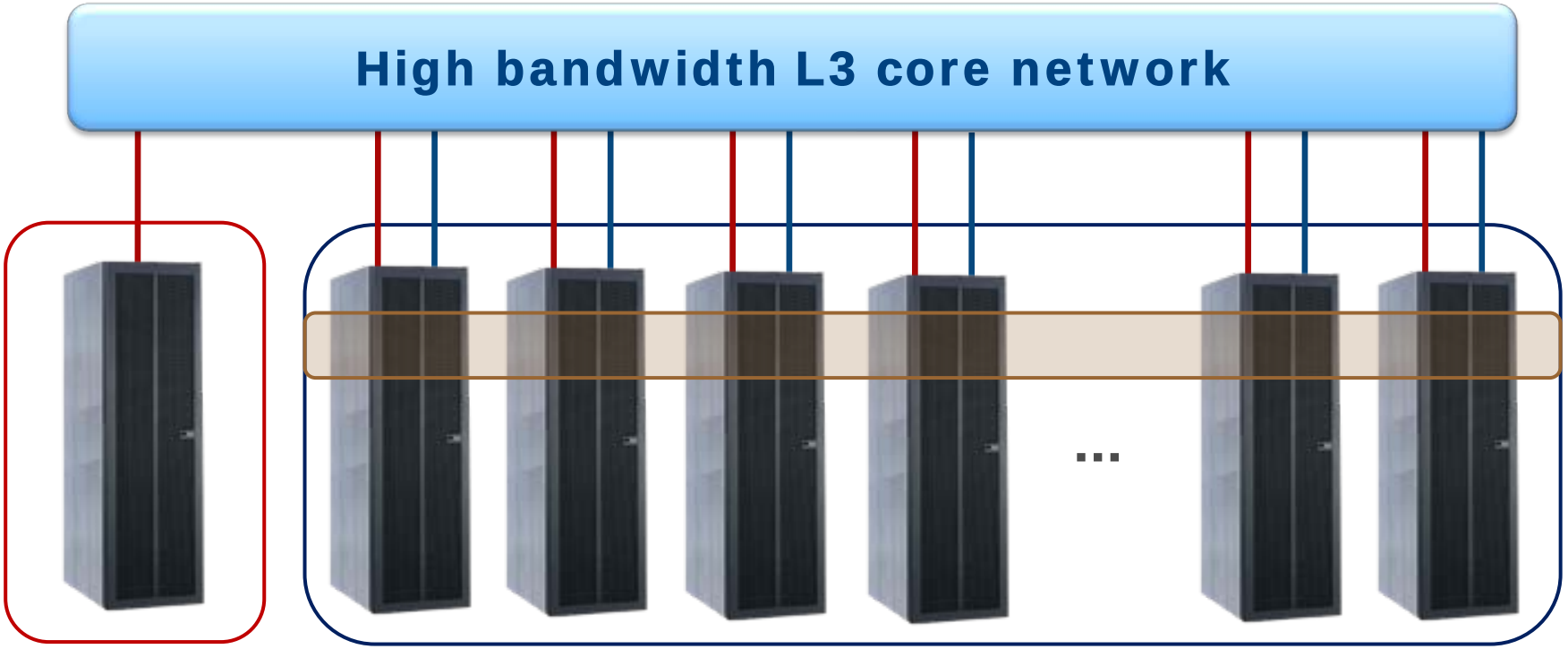
64 Microservers using Intel

Note: 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

Agenda

- Data Center Trends
- New Microserver Solutions
- Intel® Ethernet Switch Silicon Architecture
- Rack Scale Architecture
- Software-defined Infrastructure
- Summary

Transformation of the Data Center



SDN Controller

- Scalable SDN solution
- Open software interface
- Orchestrates network overlays
- Resides on VMs

Virtualized Servers and Storage

- Flexible, dense compute nodes
  - Flexibility networking
- SDN enabled
- Network Overlay tunneling
  - VxLAN, NVGRE, EVB
- Server load balancing / NAT

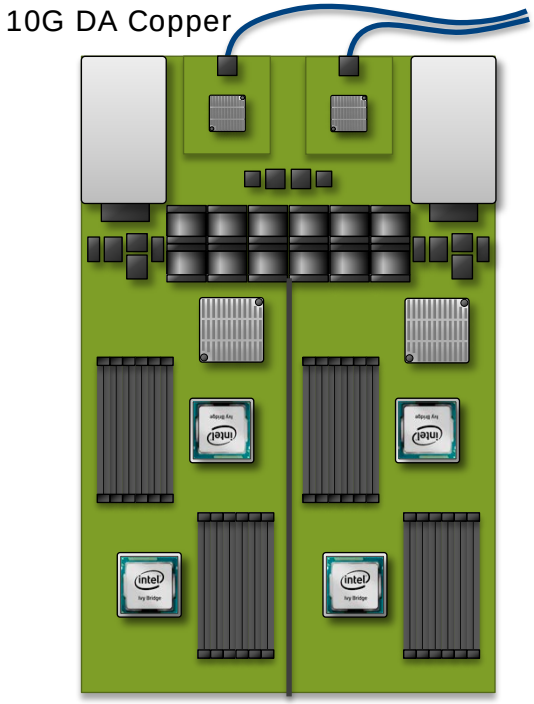
Network Function Virtualization

- Using existing infrastructure
- High bandwidth flow processing using DPDK
- Network monitoring and security
- Load balancing
- WAN optimization

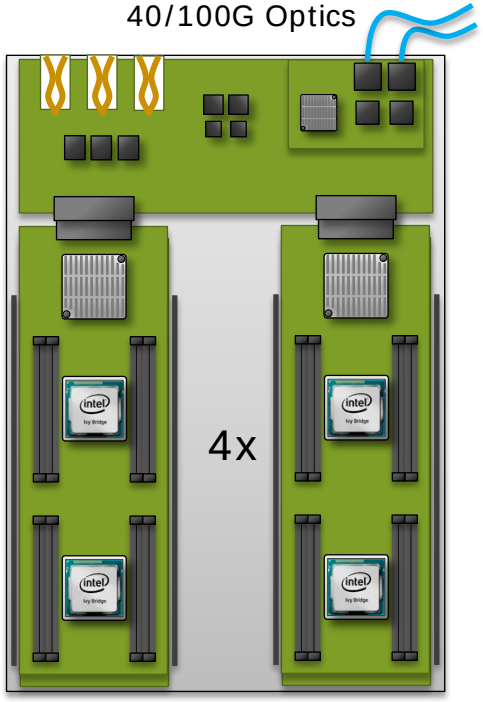
The rack becomes the basic building block

Transformation of the Server Shelf

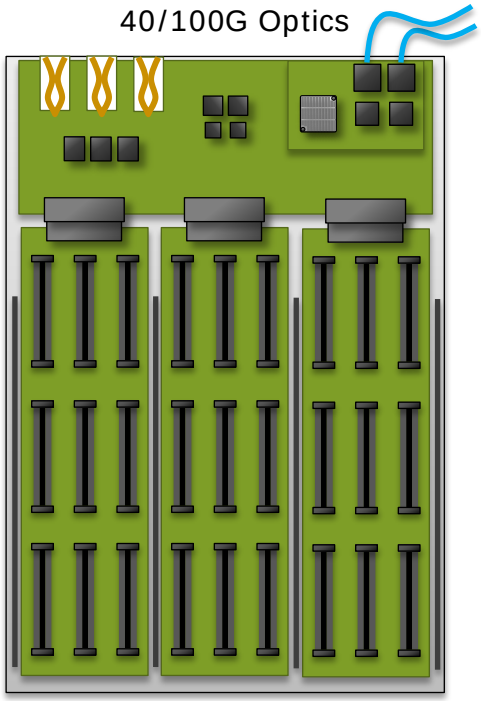
Traditional Intel® Xeon® LOM



RSA Intel Xeon Fabric



RSA Intel® Atom™ processor based Microserver Fabric



Dual LOM
- Single host per LOM
- 10GbE DA Copper
- Connected with ToR switch

Single Switch Module
- Four hosts per switch
- 40/100GbE SiPh
- Ring or mesh interconnect

Single Switch Module
- 24-30 uServers per switch
- 40/100GbE SiPh
- Ring or mesh interconnect

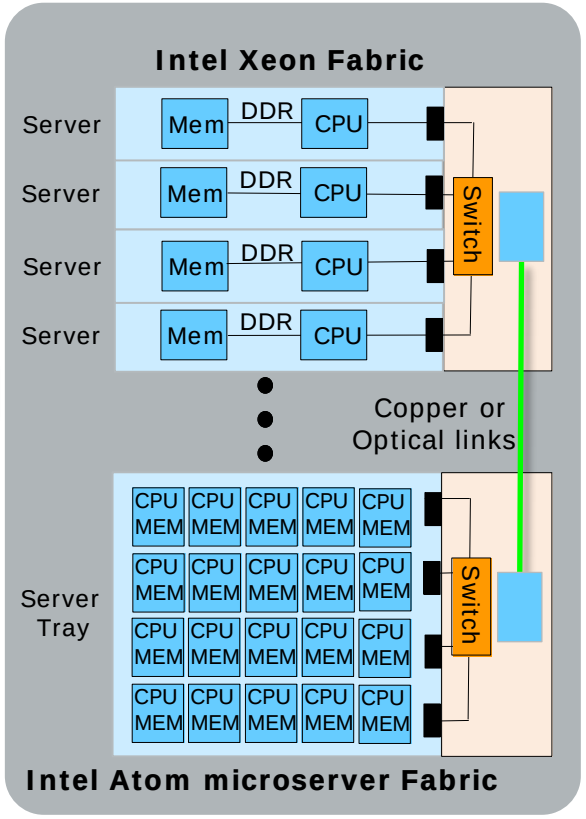
Goal: Refresh the CPUs independent of everything else

New Rack Scale Architecture Solutions

Flexible Resource Allocation
- Intel® Xeon® processor based server trays
- Microserver trays
- RAS Storage
- Flexible fabric config

Cable Reduction
- 40/100G links
- Multiple servers per link
- Improved bandwidth per server
- Manages fabric sprawl

Improved Server Density
- 4 Intel Xeon processors per shelf
- 28 Intel® Atom™ processors per shelf
- Plus lower latency



Efficient Rack Scale Server Connectivity

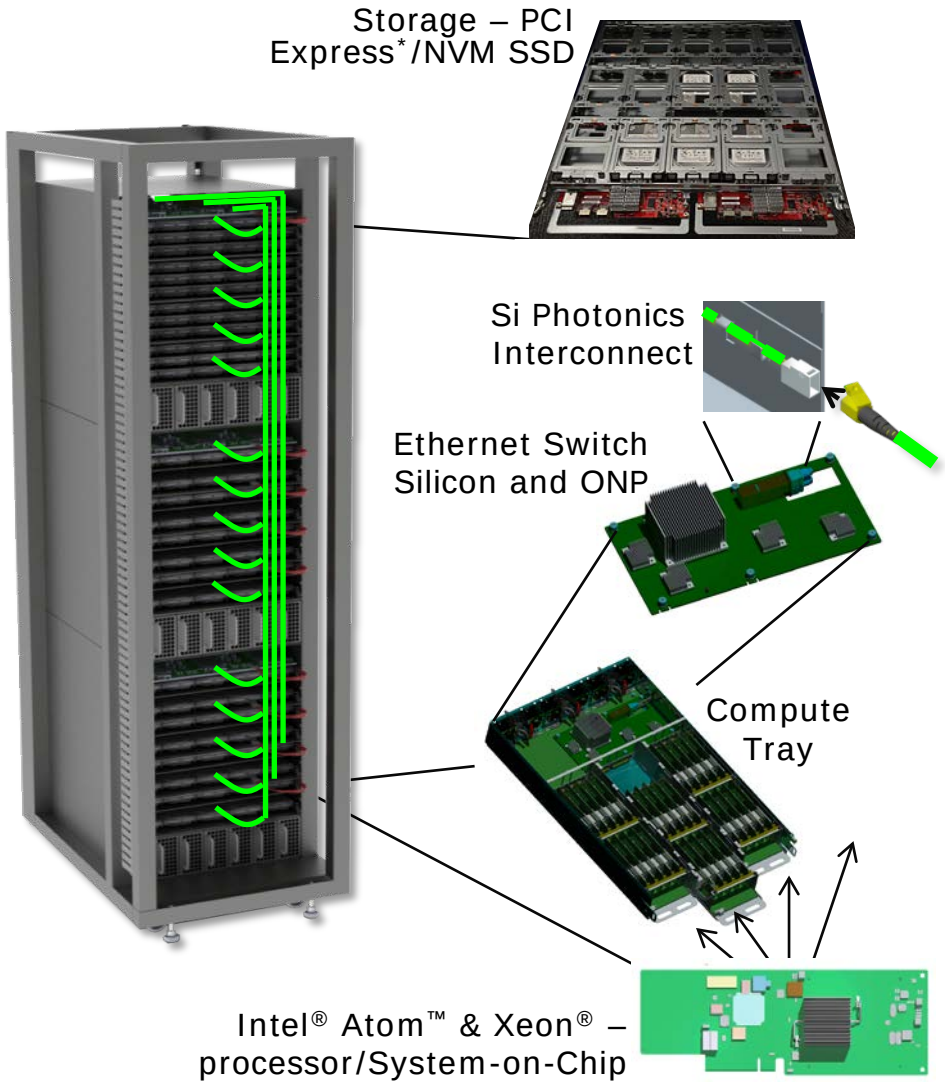
Rack Scale Architecture Components

Business Challenges

- Accelerate time to market to add capacity
- Flexible architecture
- Increase storage utilization through aggregation
- Increase platform life span, performance and capacity

Usage & Features

- Private & public cloud
- Subsystem separation
- Density/power/ performance optimized
- Open Network Platform
- Improved CapEx and OpEx



Agenda

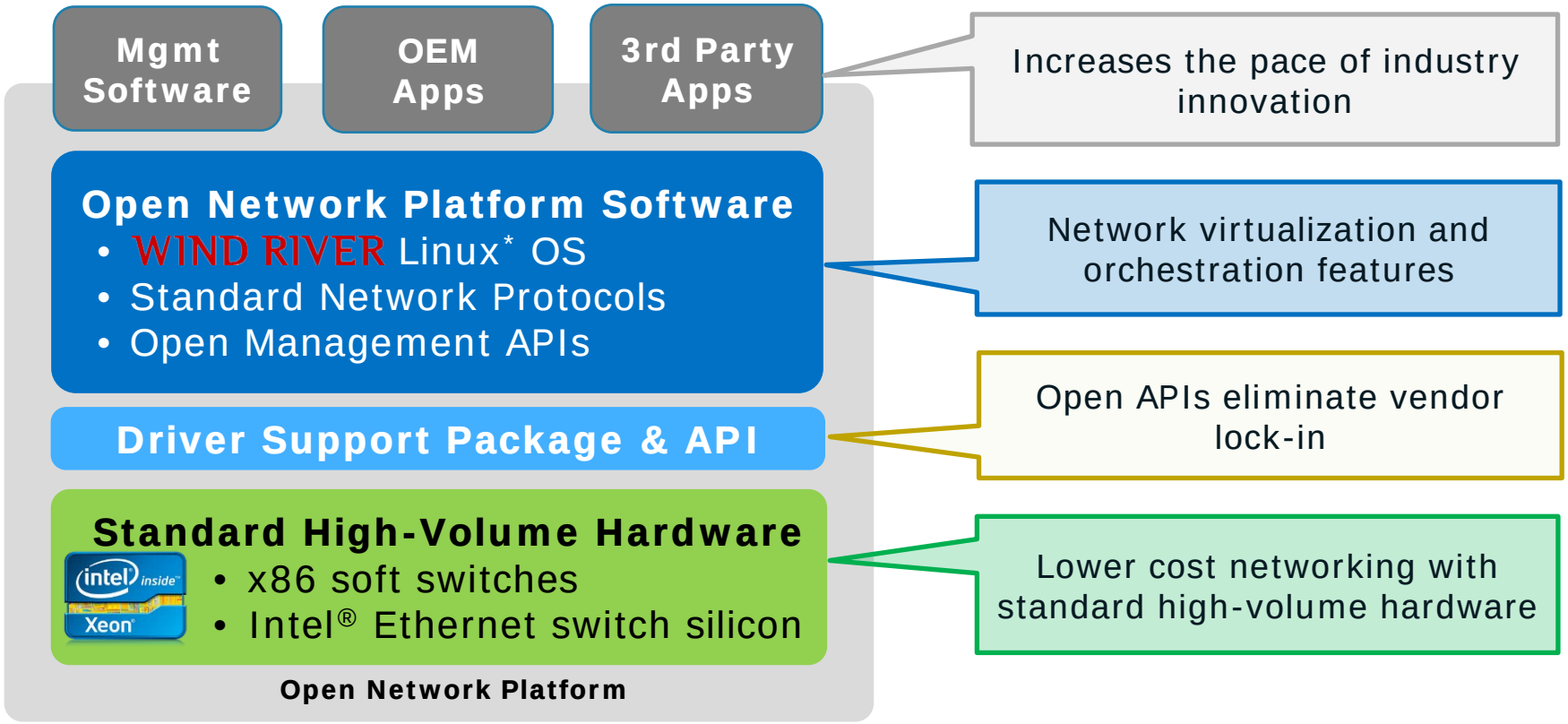
- Data Center Trends
- New Microserver Solutions
- Intel® Ethernet Switch Silicon Architecture
- Rack Scale Architecture
- Software-defined Infrastructure
- Summary

Proprietary Networking Software is not Cost Effective

- Traditional data center networking software
  - Tied to OEM hardware solutions: customer lock-in
  - Feature set mismatch to the cloud
  - Slower innovation pace
  - More costly
  - Separate software for different equipment
- The end result
  - Higher CapEx and OpEx
  - Less efficient network orchestration
  - New features are developed more slowly

Open Network Platform (ONP)

Open platform for software-defined networking




Standard CPUs


NIC Silicon


Chipset Acceleration


Open Software

Intel® Ethernet Switch FM5224

- 2.5G/10G/40G Microserver switch
- SDN-enabled switch silicon
- Intel® Flexpipe™ Technology frame processing

IDF13

Intel® Open Network Platform (Intel® ONP) Evolution

Key Ingredients

- **Wind River Open Network Software**
- **Intel® Architecture Processors**
- **Intel® Ethernet Switch Silicon**

Intel® Open Network Platform (Intel® ONP) Switch Reference Design (Seacliff Trail)

- Intel® Ethernet Switch FM6764 silicon
- High Bandwidth Low Latency
- SDN features
- ONS version 1.0

Intel ONP Rack Scale Architecture Switch Module PoC

- Passage Key
- Intel® Ethernet Switch FM6364 silicon
- Intel® 82599Ex 10 Gigabit Ethernet Controller (formerly codename Niantic)

Quanta* IZ1 ONP Productions Switch

- Based on Seacliff Trail

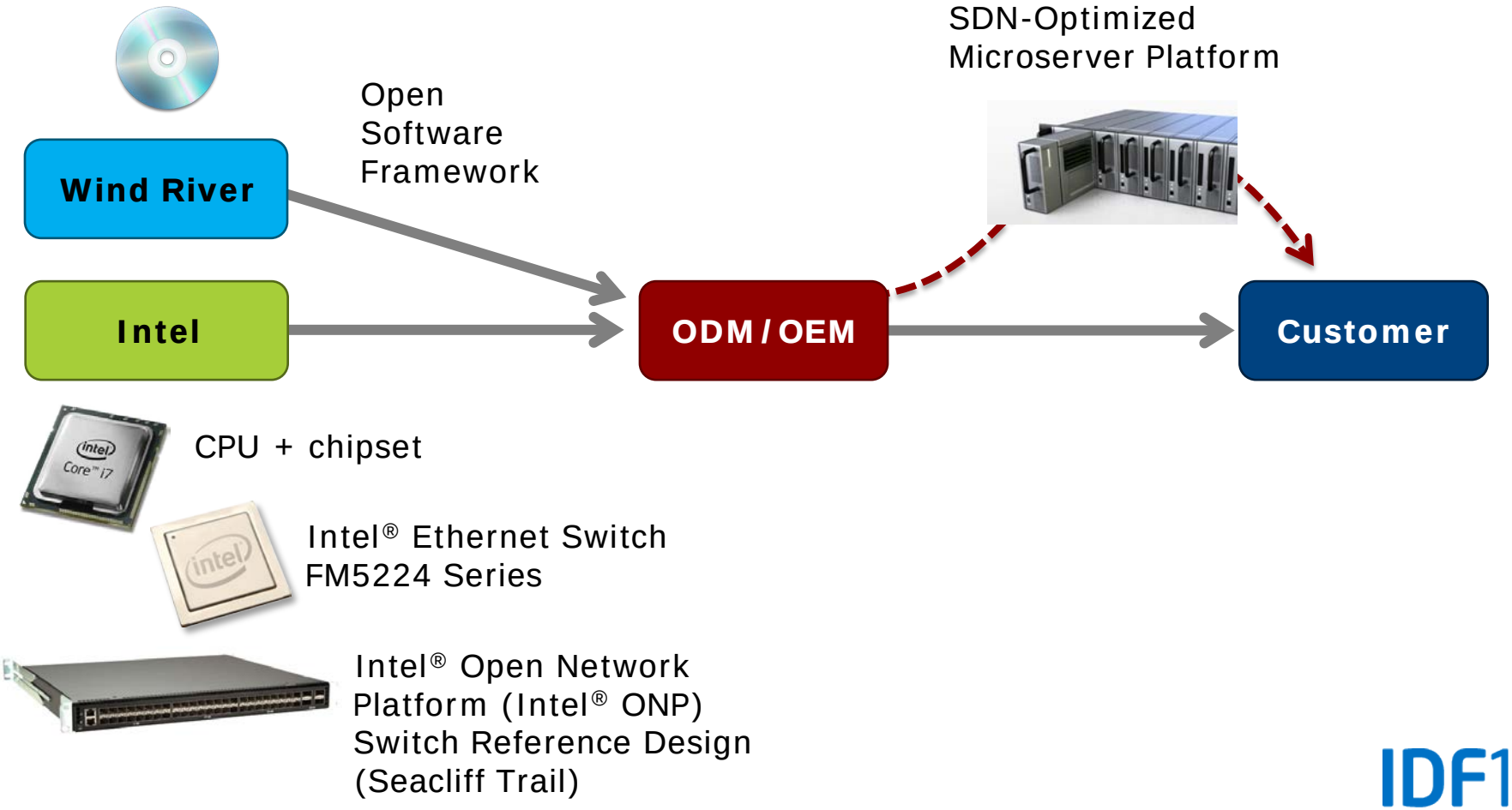
Intel ONP Microserver Switch Module

- Intel® Ethernet Switch FM5224 silicon
- Intel® processors for the networking and communications market, codename Rangeley
- ONS version 1.2

IDF13

Intel® Open Network Platform (Intel® ONP) Provides Fast SDN Deployment

- Intel and Wind River provide the basic building blocks
- Fast time-to-market using reference hardware and software



Summary

- New high density data centers lead to changes in the network
- Customers need to think about the total system
- Intel® Ethernet switch silicon and Open Network Platform (Intel® ONP) offer solutions for ToR and microservers today, RSA in the future
- Intel offers an end-to-end solution for the latest server architectures

Call to Action

- Determine your microserver and rack scale architecture requirements
- Estimate your CapEx and OpEx cost saving by moving to a software-defined infrastructure solution
- Visit the links on the next slide for more information

Additional Sources of Information

PDF of this presentation is available is available from our Technical Session Catalog: www.intel.com/idfsessionsSF. The URL is on top of Session Agenda Pages in Pocket Guide.

- For more information, come and visit us at the Software Defined Infrastructure and Data Center Community in the Technology Showcase
- Info on the Intel® Ethernet switch FM5224: http://www.intel.com/content/www/us/en/switch-silicon/ethernet-switch-fm5224-series.html
- Info on the Intel® microserver components: http://www.intel.com/content/www/us/en/servers/microservers.html
- Info on the Intel® silicon photonics: https://www-ssl.intel.com/content/www/us/en/research/intel-labs-silicon-photonics-research.html
- Article on the Rack scale architecture from Intel: http://www.theregister.co.uk/2013/07/29/intel_data_center_rack_scale_analysis/?page=1

Legal Disclaimer

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

A "Mission Critical Application" is any application in which failure of the Intel Product could result, directly or indirectly, in personal injury or death. SHOULD YOU PURCHASE OR USE INTEL'S PRODUCTS FOR ANY SUCH MISSION CRITICAL APPLICATION, YOU SHALL INDEMNIFY AND HOLD INTEL AND ITS SUBSIDIARIES, SUBCONTRACTORS AND AFFILIATES, AND THE DIRECTORS, OFFICERS, AND EMPLOYEES OF EACH, HARMLESS AGAINST ALL CLAIMS COSTS, DAMAGES, AND EXPENSES AND REASONABLE ATTORNEYS' FEES ARISING OUT OF, DIRECTLY OR INDIRECTLY, ANY CLAIM OF PRODUCT LIABILITY, PERSONAL INJURY, OR DEATH ARISING IN ANY WAY OUT OF SUCH MISSION CRITICAL APPLICATION, WHETHER OR NOT INTEL OR ITS SUBCONTRACTOR WAS NEGLIGENT IN THE DESIGN, MANUFACTURE, OR WARNING OF THE INTEL PRODUCT OR ANY OF ITS PARTS.

Intel may make changes to specifications and product descriptions at any time, without notice. Designers must not rely on the absence or characteristics of any features or instructions marked "reserved" or "undefined". Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them. The information here is subject to change without notice. Do not finalize a design with this information.

The products described in this document may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Contact your local Intel sales office or your distributor to obtain the latest specifications and before placing your product order.

Copies of documents which have an order number and are referenced in this document, or other Intel literature, may be obtained by calling 1-800-548-4725, or go to: http://www.intel.com/design/literature.htm

Avoton, Seacliff Trail, Niantic, Rangeley and other code names featured are used internally within Intel to identify products that are in development and not yet publicly announced for release. Customers, licensees and other third parties are not authorized by Intel to use code names in advertising, promotion or marketing of any product or services and any such use of Intel's internal code names is at the sole risk of the user.

Intel, Xeon, Atom, Flexpipe, Look Inside and the Intel logo are trademarks of Intel Corporation in the United States and other countries.

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

IDF13

Legal Disclaimer

- Intel® Virtualization Technology (Intel® VT) requires a computer system with an enabled Intel® processor, BIOS, and virtual machine monitor (VMM). Functionality, performance or other benefits will vary depending on hardware and software configurations. Software applications may not be compatible with all operating systems. Consult your PC manufacturer. For more information, visit http://www.intel.com/go/virtualization.

Risk Factors

The above statements and any others in this document that refer to plans and expectations for the third quarter, the year and the future are forward-looking statements that involve a number of risks and uncertainties. Words such as “anticipates,” “expects,” “intends,” “plans,” “believes,” “seeks,” “estimates,” “may,” “will,” “should” and their variations identify forward-looking statements. Statements that refer to or are based on projections, uncertain events or assumptions also identify forward-looking statements. Many factors could affect Intel’s actual results, and variances from Intel’s current expectations regarding such factors could cause actual results to differ materially from those expressed in these forward-looking statements. Intel presently considers the following to be the important factors that could cause actual results to differ materially from the company’s expectations. Demand could be different from Intel's expectations due to factors including changes in business and economic conditions; customer acceptance of Intel’s and competitors’ products; supply constraints and other disruptions affecting customers; changes in customer order patterns including order cancellations; and changes in the level of inventory at customers. Uncertainty in global economic and financial conditions poses a risk that consumers and businesses may defer purchases in response to negative financial events, which could negatively affect product demand and other related matters. Intel operates in intensely competitive industries that are characterized by a high percentage of costs that are fixed or difficult to reduce in the short term and product demand that is highly variable and difficult to forecast. Revenue and the gross margin percentage are affected by the timing of Intel product introductions and the demand for and market acceptance of Intel's products; actions taken by Intel's competitors, including product offerings and introductions, marketing programs and pricing pressures and Intel’s response to such actions; and Intel’s ability to respond quickly to technological developments and to incorporate new features into its products. The gross margin percentage could vary significantly from expectations based on capacity utilization; variations in inventory valuation, including variations related to the timing of qualifying products for sale; changes in revenue levels; segment product mix; the timing and execution of the manufacturing ramp and associated costs; start-up costs; excess or obsolete inventory; changes in unit costs; defects or disruptions in the supply of materials or resources; product manufacturing quality/yields; and impairments of long-lived assets, including manufacturing, assembly/test and intangible assets. Intel's results could be affected by adverse economic, social, political and physical/infrastructure conditions in countries where Intel, its customers or its suppliers operate, including military conflict and other security risks, natural disasters, infrastructure disruptions, health concerns and fluctuations in currency exchange rates. Expenses, particularly certain marketing and compensation expenses, as well as restructuring and asset impairment charges, vary depending on the level of demand for Intel's products and the level of revenue and profits. Intel’s results could be affected by the timing of closing of acquisitions and divestitures. Intel's results could be affected by adverse effects associated with product defects and errata (deviations from published specifications), and by litigation or regulatory matters involving intellectual property, stockholder, consumer, antitrust, disclosure and other issues, such as the litigation and regulatory matters described in Intel's SEC reports. An unfavorable ruling could include monetary damages or an injunction prohibiting Intel from manufacturing or selling one or more products, precluding particular business practices, impacting Intel’s ability to design its products, or requiring other remedies such as compulsory licensing of intellectual property. A detailed discussion of these and other factors that could affect Intel’s results is included in Intel’s SEC filings, including the company’s most recent reports on Form 10-Q, Form 10-K and earnings release.

Rev. 7/17/13