Cisco Nexus 3000 Switch Architecture

Faraz Taifehesmatian, Technical Marketing Engineer
CCIE R&S, DC
BRKDCN-3734
Cisco Webex Teams

Questions?
Use Cisco Webex Teams (formerly Cisco Spark) to chat with the speaker after the session

How
1. Find this session in the Cisco Live Mobile App
2. Click “Join the Discussion”
3. Install Webex Teams or go directly to the team space
4. Enter messages/questions in the team space

Webex Teams will be moderated by the speaker until June 18, 2018.

cs.co/ciscolivebot#BRKDCN-3734
Session Abstract

This session presents an in-depth study of the architecture of the latest generation of Nexus 3000 top-of-rack data center switches. Topics include merchant silicon architecture and capabilities (Broadcom Trident3, Tomahawk 2, Jericho+ and Barefoot Tofino), forwarding hardware, and other physical design elements, as well as a discussion of key hardware-enabled features and capabilities that combine to provide high-performance, low latency data center network services.
What This Session Covers

• Overview of Merchant Silicon with Cisco Nexus Switches
• Latest generation of Nexus 3000 switches
• System and hardware architecture, key forwarding functions, packet walks

Not covered:
• Nexus 9000 ASIC/platform architectures
• Nexus 9500 merchant-silicon based architectures
• Other Nexus platforms
Agenda

• Merchant Silicon Overview
• Nexus 3000 Portfolio
  • ASIC overview
  • Forwarding Pipeline
  • Platform Specific Details
• Key Takeaways
Merchant Silicon Overview
Merchant Silicon in Cisco DC switches

**N3000**
- StrataXGS
  - Trident +, 480 Gbps

**N3100,N9300**
- StrataXGS
  - Trident II, 1280 Gbps

**N3100,N9500-R**
- StrataXGS
  - Tomahawk, 3.2 Tbps
  - Trident II+, 1.28 Tbps
- StrataDNX
  - Jericho, 600 Gbps

**N3200, N3100V**
- StrataXGS
  - Tomahawk II, 6.4 Tbps
  - Trident III, 3.2 Tbps
- StrataDNX
  - Jericho II, 2.4 Tbps

**N3600R**
- Barefoot
  - Tofino, 1.8, 6.4 Tbps

**N3100Z,N3200E,N3400**
- StrataXGS
  - Trident II, 1280 Gbps
- StrataDNX
  - Jericho+, 900 Gbps
Nexus 3000 Portfolio
Nexus 3000 Series Switch Portfolio

**Nexus 3100**
- **ToR Leaf**
- Full-featured DC access
- Broad switch portfolio
- Based on Trident ASIC family

**Nexus 3200**
- **Fixed High Density**
- High throughput & performance
- Flexible connectivity options
- Based on Tomahawk ASIC family

**Nexus 3400**
- **Programmable pipeline**
- Support for P4-INT
- Enable custom use cases
- Based on Tofino ASIC

**Nexus 3500**
- **Ultra Low Latency**
- Financial/HFT workloads
- Based on Cisco Monticello ASICs

**Nexus 3600**
- **Deep Buffer**
- High route scale
- Video & Drop sensitive deployments
- Based on Jericho+ ASIC family
Nexus 3100 Switch Family
• Trident 3 ASIC Architecture
• ASIC Single Pipeline Block
• N3K-C3132C-Z Switch Architecture
• N3K-C3132C-Z ASIC Port-map
Nexus 3100 Switch Family

3100
- Based on Trident 2
- 1.28Tbps
- 12.2 MB Buffer

3100-V
- Based on Trident (2+)
- 1.28Tbps
- 16MB Buffer

3100-Z
- Based on Trident (3)
- 3.2Tbps
- 32MB Buffer
Trident 3 ASIC Architecture

- BCM56870 from StrataXGS family
- 3.2Tbps Single Chip Ethernet Switch
- 2 Pipes @1.6 Tbps
- 32 MB of Buffer
Trident 3 ASIC Single Pipeline Block

- FleXGS Ingress Pipeline 0:
  - Port Macro
  - Parser
  - Lookup Engine(s)
  - Lookup Engine(s)
  - Lookup Engine(s)
  - Special Functions, ECMP, Hash

Memory Management Unit (MMU)

- FleXGS Egress Pipeline 0:
  - Port Macro
  - Editor
  - Editor Control
  - Lookup Engine(s)

Cisco Live!
Trident 3 Cut-Through Vs Store-and-Forward

- Trident 3 MMU supports both store-and-forward (SF) and cut-through (CT) modes
- In SF mode, an entire incoming packet is written into the buffer first. The packet is held in the buffer until the scheduler selects that particular egress port's queue

CT mode is used in latency-sensitive applications

- In CT mode, the packet is scheduled through the cut-through path and dequeued to the EP before it has been completely received from the ingress pipeline.

- In CT mode, After one or more packet cells are received by the MMU, the packet becomes eligible for dequeuing
### Trident 3 Cut-Through switching Matrix

<table>
<thead>
<tr>
<th>Min Ingress Speed</th>
<th>Max Egress Speed</th>
<th>Destination port Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>10G</td>
<td>50G</td>
<td>10G</td>
</tr>
<tr>
<td>25G</td>
<td>50G</td>
<td>25G</td>
</tr>
<tr>
<td>40G</td>
<td>100G</td>
<td>40G</td>
</tr>
<tr>
<td>50G</td>
<td>100G</td>
<td>50G</td>
</tr>
<tr>
<td>100G</td>
<td>100G</td>
<td>100G</td>
</tr>
</tbody>
</table>

- 10G -> 10G
- 25G -> 10G
- 40G -> 10G
- 50G -> 10G
N3K-C3132C-Z

Beacon LED
Status LED
Environment LED

Lane-selected LED

2x10GE SPF+

32 40/100GE QSFP28

Each QSFP28 port can operate at 10, 25, 40, 50, and 100 G
Up to a maximum of 128 x 25-Gbps ports

1 RU
N3K-C3132C-Z Switch Architecture

- Trident 3
  - CPU: 1.8 GHz Intel Xeon
  - x4PCIe
  - MC
  - 2 x 10 G SFP
  - 32 x QSPF28

- MI FPGA
  - CPLD

- Management Console USB
- Fan Interface Board
- Haggan Board
- Chimay CPU Board
- G1x1PCIe
- G2x4PCIe
- G1x1PCIe CTRL/Status
- Fan 1
- Fan 2
- Fan 3
- Fan 4
- Lefle Board
- 650W PSU
- 650W PSU
- Leffe Board
- Management Console
- USB
- CTRL/Status
- DIMM
- CPLD
N3K-C3132C-Z ASIC Port-map

Trident 3 FalconCores

Port Group 0
Port Group 1
Pipe 0
Pipe 1
Port Group 2
Port Group 3
Port Group 4

C3132C-Z Front Panel Ports
Nexus 3000 Series Switch Portfolio

**Nexus 3100**
- ToR Leaf
- Full-featured DC access
- Broad switch portfolio
- Based on Trident ASIC family

**Nexus 3200**
- Fixed High Density
- High throughput & performance
- Flexible connectivity options
- Based on Tomahawk ASIC family

**Nexus 3300**
- Programmable pipeline
- Support for P4-INT
- Enable custom use cases
- Based on Tofino ASIC

**Nexus 3500**
- Ultra Low Latency
- Financial/HFT workloads
- Based on Cisco Monticello ASICs

**Nexus 3600**
- Deep Buffer
- High route scale
- Video & Drop sensitive deployments
- Based on Jericho+ ASIC family
Nexus 3200

- Nexus 3200 Switch Family
- Tomahawk2 ASIC Architecture
- ASIC Forwarding Pipeline
- N3K-C3264C-E Switch Architecture
- N3K-C3264C-E ASIC Port-map
Nexus 3200 Switch Family

3200
- Based on Tomahawk
- 3.2Tbps
- 16MB (4x4MB) Buffer

3200-E
- Based on Tomahawk 2
- 6.4Tbps
- 42MB (4x10.5MB) Buffer
Tomahawk2 ASIC Architecture

- BCM56970 from StrataXGS family
- 6.4Tbps Single Chip Ethernet Switch
- 4 Pipes @1.6 Tbps
- 42MB (4x10.5MB) of Buffer
- Ingress & Egress Packet Time Stamping
Tomahawk2 ASIC Single Pipeline Block
## Tomahawk2 Cut-Through switching Matrix

<table>
<thead>
<tr>
<th>Min Ingress Speed</th>
<th>Max Egress Speed</th>
<th>Destination port Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>10G</td>
<td>50G</td>
<td>10G</td>
</tr>
<tr>
<td>25G</td>
<td>50G</td>
<td>25G</td>
</tr>
<tr>
<td>40G</td>
<td>100G</td>
<td>40G</td>
</tr>
<tr>
<td>50G</td>
<td>100G</td>
<td>50G</td>
</tr>
<tr>
<td>100G</td>
<td>100G</td>
<td>100G</td>
</tr>
</tbody>
</table>

10G -> 10G
25G -> 10G
40G -> 10G
50G -> 10G
N3K-C3264C-E

Each QSFP28 port can operate at 10, 25, 40, 50, and 100 Gbps. Up to a maximum of 128 x 25-Gbps ports.
N3K-C3264C-E Switch Architecture
N3K-C3264C-E ASIC Port-map

Tomahawk2 FalconCores

C3264C-E Front Panel Ports
N3K-C3264C-E Breakout Port-mode

Following Port-modes are supported:

- 96x50g + 16x100g
- 96x25g + 32x100g
- 128x25g

CLI commands for Port-mode change:

Profile CLI: “hardware profile portmode <config>”
<config> is “96x50g+16x100g” or “96x25g+32x100g” or “128x25g”

Dynamic breakout CLI: “interface breakout module 1 port <front_port_num> map <config>”
<config> is “50g-2x” or “25g-4x” or “10g-4x”

System reload is required
N3K-C3264C-E 96x50g+16x100g

- All ports are operational and first 48 front ports will support 2x50G dynamic breakout.
- 64x100G, 64x50G + 32x100G, 96x50G + 16x100G will be met with this h/w profile
- 49-64 MACSEC Ports will support 100G and 40G operational modes
- SLIC adaptor is not supported in this h/w profile

[Diagram showing breakout capability.]
#CLUS

N3K-C3264C-E 96x25g+32x100g

- Front port 1-24, 29-32, 37-64 will be operational
- 1-24 front ports will support 2x50G, 4x25G, 4x10G dynamic breakout
- 29-32, 37-48, 49-64 MACSEC Ports will support 100G and 40G operational modes
- SLIC adaptor is supported on 1-24 Front port

<table>
<thead>
<tr>
<th>Port</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-24</td>
<td>Breakout Capable port</td>
</tr>
<tr>
<td>25-64</td>
<td>No Breakout</td>
</tr>
<tr>
<td>65-96</td>
<td>Disabled</td>
</tr>
</tbody>
</table>

![Port Breakout Diagram]
N3K-C3264C-E 128x25g

- Front port 1-28, 33-36 will be operational
- 1-28, 33-36 front ports will support 2x50G, 4x25G, 4x10G dynamic breakout
- SLIC adaptor is supported on 1-24 Front port

<table>
<thead>
<tr>
<th>Breakout Capable port</th>
<th>Disabled</th>
</tr>
</thead>
</table>

![Port Breakout Diagram]
Nexus 3000 Series Switch Portfolio

Nexus 3100
- ToR Leaf
- Full-featured DC access
- Broad switch portfolio
- Based on Trident ASIC family

Nexus 3200
- Fixed High Density
- High throughput & performance
- Flexible connectivity options
- Based on Tomahawk ASIC family

Nexus 3400
- Programmable pipeline
- Support for P4-INT
- Enable custom use cases
- Based on Tofino ASIC

Nexus 3500
- Ultra Low Latency
- Financial/HFT workloads
- Based on Cisco Monticello ASICs

Nexus 3600
- Deep Buffer
- High route scale
- Video & Drop sensitive deployments
- Based on Jericho+ ASIC family
Nexus 3400

- Barefoot Tofino ASIC Architecture
- Tofino Simplified Block Diagram
- Programmable Switch Approach
- Match-Action Packet Processing
- Match-Action Unit
- Hardware Telemetry
- Nexus 34180YC Switch Architecture
- Nexus 34180YC ASIC Port-map
Barefoot Tofino ASIC Architecture

- BFN-T10-018D from Tofino family
- 1.8Tbps Single Chip Ethernet Switch
- 2 Pipes @0.9 Tbps
- P4-programmable pipeline
- Single 20 MB Unified Packet Buffer
- Inband Network Telemetry (INT)
Tofino Simplified Block Diagram

- **Reset and Clocks**
- **PCIe**
- **CPU MAC**
- **DMA Engines**

**Control and Configurations**

**Pipe0**
- 10/25/40/50/100G Rx MACs
- Ingress Match-Action Pipeline
- Common Queuing and Packet Data Buffers
- Egress Match-Action Pipeline
- 10/25/40/50/100G Tx MACs

**Pipe1**
- 10/25/40/50/100G Rx MACs
- Ingress Match-Action Pipeline
- Common Queuing and Packet Data Buffers
- Egress Match-Action Pipeline
- 10/25/40/50/100G Tx MACs
Tofino Programmable Switch Approach

Bottom-up Network element design

Network Demands → Switch OS → Run Time API → Driver → ASIC → Fixed-function Switch

Top-down Network element design

Network Demands → NX-OS → Auto Generated API → Driver → Tofino → Programmable Switch
What Is P4?

- P4 – Programming Protocol-Independent Packet Processors
- Programming language designed to allow the definition of data planes
- Open-source, permissively-licensed language
- Designed to be protocol-independent, implementation-independent
- Protocol independence and the abstract language model allow for re-configurability, target-independence
Tofino Match-Action Packet Processing

Programmable Ingress Match-Action Pipeline

Programmable Egress Match-Action Pipeline

Programmable Parser

Buffer

PHV

Match Table

Parameters

ALUs

Action

(Match Action Unit 0)

PHV

(PHV out)

(PHV in)

PHV' (Match Action Unit N)

PHV

Parameters

ALUs

Action

(PHV out)

(PHV in)
PISA: Protocol Independent Switch Architecture

Multiple simultaneous lookups and actions can be supported.

Match + Action Stage (Unit)

N lookups → M actions
PISA: Match and Action are Separate Phases

Sequential Execution (Match dependency)

Total Latency = 3
PISA: Match and Action are Separate Phases

Staggered Execution (Action Dependency)

Total Latency = 2
PISA: Match and Action are Separate Phases

Parallel Execution (No Dependencies)

Total Latency = 1.1
Tofino Ingress Processing

- All packets processed by the ingress buffer & parser
  - Parser splits packet header into separate PHV filed, TPHV files and packet body
  - PHV files traverse through Ingress Match-Action Pipeline for table lookup and manipulation
  - Deparser reassembles packets based on files in PHV
Tofino Programmable Parser

- Will receive the packet data from the Ethernet MACs and then it would parse the packet stream according to the pre-computed parse graph.

- Next, the fields from the parsed protocol headers are extracted into the corresponding PHVs.

- Once the Parser has assembled a PHV it can then insert that PHV into the Match-Action Pipeline.

A single Parser unit can process packets at about 100Gb/s, it connects to either:
- 4 x 10/25Gb/s MACs,
- 2 x 40/50Gb/s MACs, or
- 1 x 100Gb/s MAC.
Tofino Egress Pipeline

• Egress parser extracts metadata from ingress and packet header from the packet
• Egress Match-Action Pipeline performs additional processing
• Egress deparser assembles outgoing packet
Tofino Egress Match-Action Pipeline

• Additional lookups for packet header modifications (i.e. tunnel encap, multicast replicated packets)

• Perform calculations (such as WRED) based on intrinsic metadata from TM

• Additional stats and policing as specified by P4 program
Tofino Egress Deparser

- PHV data is reassembled with packet payload
- Unnecessary fields are omitted from reassembled packet
- Final outgoing packet length fed back to TM for scheduling and shaping feedback
- Optionally send copy of packet to mirror buffer for egress mirroring
- Optionally capture PHV data into digest buffer for coalescing

Fields Dictionary

<table>
<thead>
<tr>
<th>Eth(L2)</th>
<th>Vlan</th>
<th>IPv4</th>
<th>TCP</th>
<th>UDP</th>
</tr>
</thead>
</table>

All fields with holes removed
Tofino Combined Ingress/Egress Pipeline

Combined Ingress/Egress Match-Action Pipeline

Ingress Buffer

Egress Parser 0

Egress Parser m

Ingress Parser 0

Ingress Parser m

Recirculation Buffer

Ingress Packet Body

Egress Packet Body

MAU 0

PHV

MAU n

PHV

Ingress Deparser

Egress Deparser

Ingress Packet Constructor

Egress Packet Constructor

Queues And Packet Buffers

MAC

MAC

MAC

MAC

MAC

MAC
Telemetry Modes

Postcard Mode

- In the postcard mode, each network device generates its own telemetry reports
- The collector will receive reports from different network devices, each describing the telemetry metadata

Inband Network Telemetry (INT)

- Metadata is embedded in between the original headers of data packets as they traverse the network
- This is done by INT data plane specifications
Postcard Mode vs INT Mode

Switch A
Switch B
Switch C
Collector

Event detection
Watchlist
Packet Header + Info A
Packet Header + Info B
Packet Header + Info C

Endpoint (Source)
Transit
Endpoint (Sink)

Host
Host
Host

Header
Payload
Header
Payload
Header
Payload

Packet Header + Info A, B, C
Inband Network Telemetry (INT)

- First Record (INT instruction + metadata) will be inserted in data packet at **INT Source** node
- Second Record (INT metadata) will be appended to same data packet at **INT Transit** node
- Third Record (INT metadata) will be appended to INT stack at INT Sink.

**INT Sink** will remove INT record and forward to **INT Collector** while original packet will be forwarded to server facing port
**INT Per-switch information captured**

<table>
<thead>
<tr>
<th>Flow Watch List (zoom-in view per 5-tuple of flow + DSCP bits) – 1K</th>
<th>Flow Drop List (Drop due to various drop reasons) – 256</th>
</tr>
</thead>
<tbody>
<tr>
<td>Switch ID</td>
<td>Switch ID</td>
</tr>
<tr>
<td>Hop latency</td>
<td>Ingress Port ID</td>
</tr>
<tr>
<td>Queue ID + Queue occupancy</td>
<td>Egress Port ID</td>
</tr>
<tr>
<td>Ingress timestamp</td>
<td>Queue ID</td>
</tr>
<tr>
<td>Egress timestamp</td>
<td>Drop Reason</td>
</tr>
</tbody>
</table>

- **Node-to-Node**: Reserved DSCP bit will be inserted temporarily in data packets to indicate that packets also carry INT data
- **Node-to-Collector**: A UDP encapsulation is used to pack collected INT stack at INT Sink and send to collector. Flow-affinity is maintained to send same flow-record to same collector for easy processing
# NX-OS INT Configuration Model

- **Where to send?** Exporter (IP-Address, Source)
- **What to capture?** INT Record (timestamp, Queue-id)
- **Which flow?** Watch List (5-tuple, Permit/Deny)
- **What queue thresholds?** Queue Profile (Depth, Latency)
- **How to identify the flow?** Flow Profile(DSCP, Age, Latency Quantization)
- **Which drops to report?** Drop List

**Monitor Session <n>** (Role Transit/Endpoint, Mode INT/Postcard)

---

**Enable INT “feature hw_telemetry”**
NX-OS INT Configuration Example

feature hw_telemetry

inband-telemetry exporter E1
destination 10.200.20.2
source Ethernet1/10

inband-telemetry record R1
collect switch-id
collect port-id
collect queue-occupancy
collect ingress-timestamp
collect egress-timestamp

inband-telemetry watchlist ip WL1
10 permit ip 1.1.1.1/24 10.10.10.10/24
20 deny ip 2.2.2.2/24 4.4.4.4/24

inband-telemetry monitor M1
record R1
exporter E1
watchlist WL1

inband-telemetry queue-profile QP1
depth 1000
latency 1000

inband-telemetry flow-profile FP1
dscp 1
age 5
latency quantization 10

inband-telemetry system monitor M1
N3K-C34180YC

Beacon LED
Status LED
Environment LED

48 x 10/25GE SFP28

6 QSFP28 40/100GE

1 RU
Nexus 34180YC Switch Architecture

- **CPU**: 1.8 GHz Intel Xeon
- **Barefoot Tofino**
- **48 x SPF+**: 10/25G
- **MI FPGA**
- **3 x BearValley**
- **6 x QSFP+** / QSFP28: 40/100G
- **Fan Interface Board**
- **Management Console USB**: Leffe Board
- **Fan 1, Fan 2, Fan 3, Fan 4**
- **Chimay CPU Board**: G1x1PCIe, G2x1PCIe
- **650W PSU**
- **Boddington Board**: Ctrl/Status, DIMM, CPLD

CISCO SYSTEMS CONFIDENTIAL

© 2018 Cisco and/or its affiliates. All rights reserved. Cisco Public
Nexus 34180YC ASIC Port-map

Tofino ETH ports

Pipe 0
Pipe 1

34180YC Front panel ports

48 SPF28 10/25Gig
6 QSPF28 40/100Gig

3 X BearValley
Nexus 34180YC Generic Profile
48x25G, 6x100G (Breakout 100/25, 40/10)

L2

- Interfaces (access, Trunk, Q-in-Q, Port-Channels (128))
- 4k VLAN, STP, Storm Control
- Unicast Bridging (32k MAC), Multicast forwarding/ IGMP snooping
- Peer-link less VPC
- LACP/UDLD

ACL, QoS

- Ingress ACL (MACL, VACL, RACL) (7k)
- Egress Policing
- Ingress QOS (Classification, Policing, Marking, Shaping, scheduling)
- CoPP, Custom CoPP
- PFC, LLFC, ECN

L3

- Interfaces (L3, SVI (2k), L3 port-channels)
- L3 Routing v4/v6 (v4 Host 32k and v4 LPM 4k, v6 hosts 16k, v6 LPM 4k, Next hops 48k (shared))
  - ECMP (32-way, 1k groups)
    - BGP, OSPF, BFD
    - HSRP, VRRP
- Multicast Routing, PIM-SM, SSM (SG 8k/2k, *.G 4k/1k.

Data plane Telemetry

- In-band Telemetry (Flow reports, Queue reports, Drop reports) -1k Flow watchlist, 256 Drop watchlist
- SPAN/E Data plane Telemetry RSPAN/Mirroring (64 sessions)
  - Object Models
  - PTP (Boundary mode)
  - DHCP v4/v6 Relay
N3K-C34180YC L3-Heavy Profile
48x25G, 6x100G (Breakout 100/25, 40/10)

L2
- Interfaces (access, Trunk, Port-Channels (128))
- 4k VLAN, STP, Storm Control
- Unicast Bridging (2k MAC)
- Peer-link less VPC (FCS-only)
- LACP/UDLD

ACL, QoS
- Ingress ACL (RACL) (1k), (+ Minimum Ingress ACL support for System ACL)
- Egress Shaping
- QoS (Classification, Policing, Marking, Shaping, scheduling)
- CoPP, Custom CoPP
- ECN

L3
- Interfaces (L3, SVI (2k), L3 port-channels)
- L3 Routing v4 (Host 64K and LPM 64K, next hops 64K)
  - ECMP (32-way, 1k groups)
  - BGP, OSPF, BFD
  - HSRP, VRRP

Data plane Telemetry
- In-band Telemetry (Flow reports, Queue reports, Drop reports) - 1k Flow watchlist, 256 Drop watchlist
- SPAN/ERSPAN/Mirroring (64 sessions)
  - Object Models
  - PTP (Boundary mode)
  - DHCP v4 Relay
Nexus 3000 Series Switch Portfolio

Nexus 3100
- ToR Leaf
- Full-featured DC access
- Broad switch portfolio
- Based on Trident ASIC family

Nexus 3200
- Fixed High Density
- High throughput & performance
- Flexible connectivity options
- Based on Tomahawk ASIC family

Nexus 3400
- Programmable pipeline
- Support for P4-INT
- Enable custom use cases
- Based on Tofino ASIC

Nexus 3500
- Ultra Low Latency
- Financial/HFT workloads
- Based on Cisco Monticello ASICs

Nexus 3600
- Deep Buffer
- High route scale
- Video & Drop sensitive deployments
- Based on Jericho+ ASIC family
Nexus 3500

• Monticello Architecture
• ASIC Block Diagram
• ASIC Forwarding Paths
• ASIC Forwarding Pipeline
  • Normal Mode
  • Warp Mode
• Monticello Warp Span
• N3K-C3548P-XL Switch Architecture
• N3K-C3548P-XL ASIC Port-map
Monticello ASIC Architecture

- 480 Gbps Single Chip Ethernet Switch
- 720 MPPS @ 64 Bytes
- 18 MB of Buffer (3 x 6MB)
- UUL 250ns (~200ns in Warp Mode)
Monticello ASIC Block Diagram

Ports:
9-12
21-24
33-36
45-48

x16

Parser

Decision Engine

Ports:
5-8
17-20
29-32
41-44

x16

Parser

Ports:
1-4
13-16
25-28
37-40

x16

Parser

Ports:
9-12
21-24
33-36
45-48

ReWrite

Ports:
5-8
17-20
29-32
41-44

x16

ReWrite

Ports:
1-4
13-16
25-28
37-40

x16

Output Buffer 0
6MB

Output Buffer 1
6MB

Output Buffer 2
6MB

x16

Parser

Admission Control

Output Buffer 0
6MB

Output Buffer 1
6MB

Output Buffer 2
6MB

x16
Monticello ASIC Forwarding Paths

1. Classification
2. ACL
3. Normal L2
4. L2 + L3
5. L3
6. Egress Port
7. Incoming Packet
8. Warp Span
9. Warp
Monticello ASIC Forwarding Pipeline
Normal Mode
Monticello ASIC Forwarding Pipeline

WARP Mode

Parsed/Classified Traffic

Lookup Engine

Unicast LPM 4K
Multicast Table 8K
MAC Table 8K
Host Table 8K
Control 4K
Unicast Host Table 64K
ACL Table 4K

Ingress ACL Result

L3 UC ECMP Table 16K
MAC Table 64K
Decision Engine

To Buffer
## Normal vs. Warp Mode Forwarding

<table>
<thead>
<tr>
<th>Feature</th>
<th>Normal</th>
<th>Warp</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Latency</strong></td>
<td>250ns</td>
<td>190ns</td>
</tr>
<tr>
<td><strong>NAT</strong></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Ingress RACL/VACL</strong></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Egress RACL/VACL</strong></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>Unicast Route</strong></td>
<td>24K</td>
<td>4K</td>
</tr>
<tr>
<td><strong>Multicast Route</strong></td>
<td>8K</td>
<td>8K</td>
</tr>
<tr>
<td><strong>L3 ECMP</strong></td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>
Monticello Warp Span

- WARP SPAN can be enabled both in normal and WARP mode
- Latency ~50 ns
- WARP SPAN source has to be port 1/36
- Destination ports would be group of 4 ports
N3K-C3548P-XL

- 48 1/10GE SPF+
- 40 GE by combining four sequential SFP+
- 1 RU
- Same Physical device as 3524-XL
- Two USB ports
- 10/100/1000 management port
- RS-232 serial console port
N3K-C3548P-XL Switch Architecture

- CPU: 2.50GHz GHz Intel Core(TM) i3-3227U
- PCIe: 48 x SPF+ 10G MI FPGA
- CPLD: 16GB Bootflash
- DIMM: Fan Board
- Fan1, Fan2, Fan3, Fan4
- Monticello CR
- XFI: 48 x XFI
- 24 x XFI to SFI
- 48 x SFI
- Management Console USB
- 400W PSU
- 400W PSU
N3K-C3548P-XL ASIC Port-map

MonticelloCR OB Ports

48 SPF+ 10Gig
Nexus 3000 Series Switch Portfolio

**Nexus 3100**
- ToR Leaf
- Full-featured DC access
- Broad switch portfolio
- Based on Trident ASIC family

**Nexus 3200**
- Fixed High Density
- High throughput & performance
- Flexible connectivity options
- Based on Tomahawk ASIC family

**Nexus 3400**
- Programmable pipeline
- Support for P4-INT
- Enable custom use cases
- Based on Tofino ASIC

**Nexus 3500**
- Ultra Low Latency
- Financial/HFT workloads
- Based on Cisco Monticello ASICs

**Nexus 3600**
- Deep Buffer
- High route scale
- Video & Drop sensitive deployments
- Based on Jericho+ ASIC family
Nexus 3600

- Nexus 3600 Switch Family
- Jericho+ ASIC Architecture
- ASIC Packet Forwarding
- N3K-C3636C-R Switch Architecture
- N3K-C3636C-R ASIC Port-map
Nexus 3600 Switch Family

36180YC-R
- Based on 2 Jericho +
- 1.8Tbps
- 8 GB of Buffer

3636C-R
- Based on 4 Jericho +
- 3.6Tbps
- 16 GB of Buffer
Jericho+ ASIC Architecture

- BCM88680 from StrataDNX family
- 900Gbps, 835Mpps
- Integrated Forwarding and Fabric interface
- Two packet processing cores (PP)
- 96K Virtual Output Queues
Jericho+ High-Level Forwarding Architecture

On-chip resources
- 16MB Internal Buffer & TCAM
- Forwarding Tables

Expansion via off-chip resources
- Deep GDDR5 external packet buffers

Ingress/Egress Traffic Managers
- 96k Virtual Output Queues
- WRED, Distributed Arbitration
Jericho+ Buffering

- Nexus N3600-R switches use traditional VoQ architecture
- Big buffer on Ingress side dedicated to VoQ buffer
- 4GB GDDR5 DRAM-based buffering per port-group used for VoQ buffer
- VOQ buffer has dedicated portion per port and shard buffer among ports in the same port group
- 16MB of On-chip buffer used for egress buffer
N3K-C3636C-R

- RS-232 serial console port
- RS-FEC for 25GE
- 8 QSFP28 + MACsec
- 1 RU
- USB Port
- 36 QSFP28
- Each QSFP28 can breakout to 4 x 10 or 4 x 25GE
- 10/100/1000-Mbps management port (copper or fiber)
- Lane-selected LED

Each QSFP28 can breakout to 4 x 10 or 4 x 25GE
N3K-C3636C-R Switch Architecture
N3K-C3636C-R ASIC Port-map

Jericho+ Ports

C3636C-R Front Panel Ports

Jericho+ 1
Jericho+ 2
Jericho+ 3
Jericho+ 4
Nexus 3000 Series Switch Portfolio

Nexus 3100
- ToR Leaf
- Ultra Low Latency
- Full-featured DC access
- Financial/HFT workloads
- Based on Trident ASIC family
- Fixed High Density
- High throughput & performance
- Based on Cisco Monticello ASICs

Nexus 3200
- Programmable pipeline
- High route scale
- Support for P4-INT
- Video & Drop sensitive deployments
- Based on Tofino ASIC

Nexus 3400
- Programmable pipeline
- Enable custom use cases
- Based on Tofino ASIC

Nexus 3500
- Ultra Low Latency
- Fixed High Density
- Financial/HFT workloads
- Broad switch portfolio
- Based on Cisco Monticello ASICs

Nexus 3600
- Deep Buffer
- Flexible connectivity options
- High route scale
- Video & Drop sensitive deployments
- Based on Jericho+ ASIC family
Key Takeaways
Key Takeaways

Ultra-low latency
Less than 200ns
Normal, Warp, Warp SPAN

Full-featured DC access
Broad Switch Portfolio
High throughput & performance
Flexible connectivity

Programmable Pipeline
Inband Network Telemetry (INT)
In-Network DDoS Detection
Layer 4 Load Balancer

Deep Buffer
16 GB of Buffer
Ingress VoQ
Multiple ASICs

Cisco Nexus 3000 Series Switches

3100, 3200
3400
3500
3600
Complete your online session evaluation

Give us your feedback to be entered into a Daily Survey Drawing.

Complete your session surveys through the Cisco Live mobile app or on www.CiscoLive.com/us.

Don’t forget: Cisco Live sessions will be available for viewing on demand after the event at www.CiscoLive.com/Online.
Continue your education

- Demos in the Cisco campus
- Walk-in self-paced labs
- Meet the engineer 1:1 meetings
- Related sessions
Thank you