OPEN POSSIBILITIES.

Meta Next-Generation Network Switch – Minipack2 and Arista 7388X5
Meta Next-Generation Network Switch – Minipack2 and Arista 7388X5

Xu Wang, Hardware Engineer, Meta
Narayanan Suryanarayanan, Systems Engineering, Arista Networks
Agenda

• Overview
• Minipack2
• Arista 7388X5
Minipack2 & Arista 7388X5 Overview

• High Performance 25.6T switch with modular line cards
  • 8x PIM - Port Interface Module
• Based on Broadcom Tomahawk4 25.6T switch ASIC
  • 512x 53.125G PAM4 SerDes
• Supporting 128x 200G-FR4 QSFP56 optics modules
  • 4x 53.125G PAM4 host interface
• Consistent SerDes speed at the switch ASIC, the optics host interface, and on the optics line / wavelength
  • No gearbox
Minipack2 & Arista 7388X5’s Role in Meta DC Network

- F16 data center network
  - Fabric switch
  - Spine switch
Agenda

• Overview
• Minipack2
• Arista 7388X5
Minipack2 Functional Block Diagram

- SMB (1x)
- TH4 ASIC
- RunBMC
- PIM (8x)
- Retimers – BCM87326
- 16x QSFP-DD ports
- SCM (1x)
- Mini-Lake COMe
- NVMe SSD
Minipack2 Front View

- **Console**
- **Mgmt Ethernet**
- **PSU**
- **Air Channel**
- **SCM**
- **PIM2 ~ PIM9**

*OPEN POSSIBILITIES.*
Minipack2 Rear View

PSU1 ~ PSU4

SMB FRU
Release Button

Fan1 ~ Fan8
Minipack2 Orthogonal-Direct Structure

- Orthogonal-direct structure
- Opens airflow path for better thermal efficiency
- Supports 200G-FR4 optics with 65°C case temperature limit
- Shortens PCB traces for lower loss
- FRU-able, modular PIM
  - PIM-16Q
  - Easy to explore other PIM options
  - Reduced failure domain
Minipack2 Orthogonal-Direct Structure (2)
Minipack2 Assembly View

- SMB
- PIM-16Q
- SCM
- Front FRU ejectors
- SMB ejectors

Open Possibilities.
Optics Transceivers Supported

- 128x 200G-FR4 QSFP56 optics modules
  - 4x 53.125G PAM4 host interface
- 64x 400G-FR4 QSFP-DD optics modules
  - 8x 53.125G PAM4 host interface
- 128x 100G-CWDM4(-Lite) QSFP28 optics modules
  - 4x 25.78125G NRZ host interface
## System Power Consumption

<table>
<thead>
<tr>
<th>Fan PWM</th>
<th>30%</th>
<th>35%</th>
<th>40%</th>
<th>45%</th>
<th>50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inlet temperature (°C)</td>
<td>25</td>
<td>25</td>
<td>35</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>System AC Power Input (W)</td>
<td>1711</td>
<td>1713</td>
<td>1737</td>
<td>1752</td>
<td>1771</td>
</tr>
</tbody>
</table>

- **Test Conditions**
  - Full line rate L2 snake test with 295-byte packets
  - All ports were populated with 200G-FR4 QSFP56 optics.
- The average system power consumption at full line rate under typical conditions is budgeted at 1,750W.
SMB

- Broadcom TH4 25.6T switch ASIC
  - 512x 53.125G PAM4 SerDes
  - LGA socketed
- RunBMC module w/ Aspeed AST2620
- Broadcom BCM5387 management Ethernet switch (OOB)
- Microchip PM40028 PCIe switch
- IOB FPGA
- 40-layer PCB
  - Doosan DS-7409DV(N) and FaradFlex MC24M
  - 14 signal layers and 6 power planes
  - Single lamination PCB w/ PTH vias only
SMB Signal Routing

- Complex routing in order to load balance the PIM slots
RunBMC

- **Key components**
  - Aspeed AST2620
  - 2x SPI NOR boot Flash (128 MB)
  - 16 Gb DDR4 memory down
  - 8 GB eMMC
  - TPM 2.0

- **Module form factor**
  - 260-pin right-angle SO-DIMM connector
  - 69 mm x 50 mm
  - Pin-out defined by the RunBMC OCP spec, with minor customization
  - Placement on SMB
PIM-16Q

- 4x Broadcom BCM87326 retimers
  - 16x 53.125G PAM4 SerDes lanes
- 16x QSFP-DD ports
  - 4x 1x4 belly-to-belly mounted QSFP-DD cages
- DOM FPGA

OPEN POSSIBILITIES.
PIM-16Q Port Speeds Supported

• Each pair of ports on the same row, e.g., P1 and P2, can work in the following modes:
  • 400G
  • 200G, 200G
  • 100G, 100G
  • 100G, 40G
  • 40G, 100G
  • 40G, 40G

• The modes of each pair of ports can be configured individually, regardless of the modes of other port pairs.
SCM

- Mini-Lake COMe
- Intel Broadwell-DE D1527
- 2x 16GB DDR4 SO-DIMM
- NVMe SSD
- Front panel ports
  - Serial console to BMC
  - Mgmt Ethernet
IOB FPGA on SMB

- IOB FPGA has a USB2.0 interface to BMC and eight SPI buses for PIM FPGA image programming.
- Enables parallel PIM FPGA Flash programming through IOB FPGA.
- Enables live FPGA upgrade without impacting traffic.
- sLPC 32bit data access, instead of 16bit data access in Minipack.
DOM FPGA on PIM

- PIM FPGA supports independent MDIO interfaces to retimers.
- PIM FPGA supports 16 independent I2C controllers to QSFP port.
- PIM FPGA does not support DOM auto collection function (Deprecated from Minipack).
AC/DC PSU and Fan Tray

- The AC/DC PSU and the fan tray are the same as those in Minipack.

**OPEN POSSIBILITIES.**
Minipack2 Summary

• High Performance 25.6T switch with modular line cards
  • Similar system architecture as Minipack
• Adopting Broadcom Tomahawk4 25.6T switch ASIC
  • With Broadcom BCM87326 retimers on PIM
• Supporting 128x 200G-FR4 QSFP56 optics modules
  • Keeping QSFP56 case temperature under 65°C
• Enabling network upgrade from 100G to 200G
  • With the same DC infrastructure
Agenda

- Overview
- Minipack2
- Arista 7388X5
Arista 7388X5 – Co-Development

• Cooperatively developed with Meta
• Shares all the key characteristics of Minipack2
• Deliberately designed to share critical operator aspects:
  ○ Port map, access to BMC, green touch points, compatible rack kits...
• Brings benefits of hyperscale deployments to all operators
• Significantly reduces power per bit compared to Minipack1 / 7368X4
• Increase in features and table sizes to support more use cases
• Support for Arista EOS, FBOSS & SONiC (planned)

OPEN POSSIBILITIES.
Arista 7388X5 High Level Block Diagram

- **Switch Card** – 128 x 200G / 64 x 400G
- **Management Card**
- **PSU**
- **Fan**
- **QSFP56-200** (16 x 200G)
- **QSFP56-DD** (8 x 400G)
- **QSFP56-200** (16 x 200G)
7388X5 Front View – 128 x 200G

- Management Module
- Ejector tool
- Linecard Ejector
- 16 x 200G QSFP-56
- 16 x 100G QSFP
- Linecard2 ~ Linecard9

OPEN POSSIBILITIES.
7388X5 Front View – 64 x 400G

- Ejector tool
- Management Module
- 8 x 400G QSFP-DD
- 8 x 100/200G QSFP56/28
- Linecard2 ~ Linecard9
- Linecard Ejector
7388X5 Rear View

- Single Switch Card / Module
- 2U Fan Modules
- Switchcard Ejector
- AC / DC PSU 2400W (HV)
- Upto 4 PSU (2 Default)
- OCP Green for Touchpoint
7388X5 – Architected for Cloud Operators

- **High Performance 25.6Tbps Modular System 4U Optimized**
  - Switch card: Removes from rear without cable changes
  - Management Module: Remove from the front
  - Power Supplies: Rear access and hot swap
  - Fan Modules: Individually replaceable with hot swap
  - Choice of port module configurations – mix and match
    - 16 x 200G QSFP56 / 100G QSFP
    - 8 x 400G QSFP-DD
    - 8 x 400G QSFP-DD with MACSec

- **Max Power Measured**
  - Under 1.75KW for 128x200G
  - Under 3.5KW for 64x400G Macsec
7388X5 – Management

- Quad core x86 CPU with hyper-threading for high performance control plane
- 32GB DRAM for control plane scale
- RunBMC module w/ Aspeed AST2620
- 512G SSD for NOS and Logging
- Console and Management Ethernet Ports
- USB port for memory devices
- LEDs for System, Fan, PSU and Switch Card
7388X5 - Optics Transceivers Supported

• 128x 200G QSFP56 optics modules
  • 200GBASE-SR4, 200GBASE-FR4
• 64x 400G-FR4 QSFP-DD optics modules
  • 400GBASE-SR8, 400GBASE-DR4, 400GBASE-XDR4, 400GBASE-PLR4, 400GBASE-FR4, 400GBASE-LR4, 400GBASE-ZR
• 128x 100G-CWDM4 QSFP28 optics modules
  • 100GBASE-SR4, 100GBASE-PSM4, 100GBASE-CWDM4, 100GBASE-DR, 100GBASE-FR, 100GBASE-LR, 100GBASE-LR4, 100GBASE-LRL4 & more
## System Power / Cooling Consumption

<table>
<thead>
<tr>
<th>Inlet Temperature (°C)</th>
<th>25C</th>
<th>30C</th>
<th>35C</th>
<th>40C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fan PWM (%)</td>
<td>23%</td>
<td>26%</td>
<td>39%</td>
<td>62%</td>
</tr>
<tr>
<td>System AC Power Input (W)</td>
<td>1219W</td>
<td>1236W</td>
<td>1285W</td>
<td>1366W</td>
</tr>
</tbody>
</table>

### Test Conditions
- Full line rate traffic with 300-byte packets
- All 128 ports populated with 200G-FR4 optics
- Maximum optics module temperature below 65C
7388X5 - Power and Cooling

2400W AC/DC PSU - Up to 277V Input
OCP Green handles
HV AC input with SAF-D connector

2RU Fans with Green Handles
for operator touch points
Arista 7388X5 Summary

- **Hyperscale 100G/200G/400G system**
  - High Performance with 25.6Tbps and 10.6Bpps
  - Low Latency - 825ns port to port with shared 114MB Smart-buffer
  - MACSec Support for 100G/200G/400G
  - Under 10W per 200G port typical to lower TCO
- **Flexible Operating Systems**
  - Arista EOS, FBOSS or SONiC (capable)
- **Hyperscale Cloud Scalability**
  - Increased routing scale and robustness - 400K routes, 128-way ECMP
  - Dynamic Load Balancing & Dynamic Group Multipath
  - Optimized hashing and ALPM for large scale IPv4 and IPv6
Call to Action

- Timeline for Contribution Availability
  - Minipack2 – Submit specs and design package for *OCP Accepted* in Nov. 2021
  - 7388X5 – Submit for *OCP Inspired*
- Timeline for Product Availability
  - Minipack2 – Early 2022 by Celestica
  - 7388X5 – Customer Qual in early 2022
- Where to find additional information
  - Minipack2 – www.celestica.com
  - 7388X5 – www.arista.com
Thank you!