OCTEON 10 DPU Family
Accelerating the data infrastructure transformation

June 2021
Workloads are shifting to data-centric compute

AI, Networking, security, video and storage virtualization
DPU definition

A Data Processing Unit (DPU) is a compute entity that is used to move, process, secure and manage data, as it travels or while at rest, to make it available and optimized for application.

Compute

Accelerators

High speed Interfaces
OCTEON 10
Industry Firsts

- Compute leadership with Arm Neoverse N2 cores
- Based on TSMC 5nm process
- Integrated hardware ML engine
- Integrated 1 terabit switch
- VPP hardware acceleration
- Advanced inline crypto accelerators

Compute leadership with industry-leading performance per Watt
OCTEON 10 Platform
Scalable system performance

- 400G+ Datapath
- NVMe
- 20M+ IOPs
- SPECint 1000+
- 400+ Gbps IPSEC
- 400+ Gbps TLS
- 100's TOPS
- 100's
OCTEON 10 platform scalability

Edge / 5G → Enterprise & Core → Cloud

Scalable features

- **Ethernet port speed**
  - 1G → 400G

- **Datapath**
  - 50G → 400G+

- **Compute**
  - SPECint 275 → 1000+

- **Security**
  - IPSEC/SSL 50G → 400G+

Common software
OCTEON DPU platform

Software

OCTEON DPU Open Software Platform

Optimized Stacks: Networking, Storage, Security
Virtualization and Containers
Standard APIs DPDK, SPDK, VPP

Silicon

OCTEON DPU

Arm cores
Ethernet / PCIe / Memory Controllers
Software-enabled Accelerators
OCTEON 10 innovations

- 5nm TSMC process
  - Enables fanless designs
- First inline DPU ML Engine
- Hardware VPP acceleration
- Inline crypto processor
- Arm Neoverse N2 cores
  - Highest SPECint in industry
- PCIe 5.0, DDR5 support
- Integrated with 16x 50GE switch
- 56G SerDes
Best-in-class DPU inferencing
- Directly in the data pipeline
- Each ML tile contains private SRAM

Up to 100x performance vs SW
- Supports Int8, FP16

Use cases
- Threat detection
- Context-aware service delivery
- QoS
- Beamforming optimization
- Predictive maintenance
VPP hardware acceleration

Today’s scalar packet processing

Packet scheduling 1 by 1

Packet engine

Packet header lookup
Packet decision logic
Header manipulation

Hardware packet scheduler packet transmit

OCTEON 10 Vector packet processing

Vectorized

Packet header lookup
Packet decision logic
Header manipulation

Processed as vector

Hardware packet scheduler packet transmit

Up to 5X system level performance gains

© 2021 Marvell. All rights reserved.
Arm Neoverse N2 advantages

- Maximizes performance per watt
  - Same software runs faster
- 3x single-threaded performance
  - Lower application latency with 1M L2 cache
- Higher performance and scalability
  - SVE2 for hyperscan/DPI and ML support
  - Enhanced cryptography instructions
- 3x latency reduction to HW accelerators
  - Enabled by hardware scheduling attached cores
OCTEON 10 Switch integration

- 1T switch with 16 x 50GE ports
  - Support from 1GE to 100GE

- Feature support
  - 256bit MACsec
  - Network overlay for VxLAN/GRE/MPLS
  - Network analytics: sFlow, IPFIX
  - Flow aware-processing
  - Line rate telemetry
  - TSN timing

- Example use cases
  - 5G: front haul, back haul, side haul
  - Edge switching
  - Enterprise ethernet port fan out
### OCTEON DPU software platform

<table>
<thead>
<tr>
<th>Optimized Stacks</th>
<th>Networking</th>
<th>Storage</th>
<th>Security</th>
<th>Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virtualization</td>
<td>Docker / KNI</td>
<td>Hypervisor</td>
<td>vSwitch</td>
<td>CNI/CSI</td>
</tr>
<tr>
<td>Open framework</td>
<td>DPDK / VPP</td>
<td>Linux Kernel</td>
<td>OpenSSL, TLS, IPSEC</td>
<td>SPDK</td>
</tr>
</tbody>
</table>
Virtualized Accelerators

Linux OS
- Linux netdev VF driver

PCIe
- VF driver
  - Bare metal
  - Container
  - VM

VF0 (v DPU)
- Accelerators

VF2 (v DPU)
- Accelerators

VF3 (v DPU)
- Accelerators

VF4 (v DPU)
- Accelerators

OCTEON 10 DPU

KVM
4G/5G RAN Architectures

Disaggregated RAN

O-RAN/vRAN Offload

Use case
Cloud and Datacenter DPU

- Compute: 1000+ SPECint
- Ethernet ports: Up to 400GE
- Datapath: 400G+

- Storage: 20M+ IOPS
- AI/ML: 100’s TOPS
- Security: 400G+ of IPSEC and SSL
## Development Platform

### Compute
- OCTEON 10 DPU
- 24 Neoverse N2 cores
- SPECint > 800

### Accelerators
- inline ML engine
- inline IPSec
- SSL/TLS
- Vector packet processing (VPP)

### SW
- DPDK networking suite for Control, Management and Fast path stacks
- SDK with Linux kernel and user plane extensions

### Memory
- 16GB DDR5-5200 + ECC on-board memory

### I/O
- 2 x 100GbE QSFP56
- PCIe 5.0

Available in Q4
Enterprise router and firewall appliance

- Scalability from low to high end systems
- OCTEON 10 compute
  - Best single threaded performance
- Data plane acceleration supports:
  - NG Firewall, IPSec, TLS
  - L2/L3 forwarding
  - Advanced packet parsing
  - Inline hardware base AI/ML
- Support for up to 20 ethernet MACs
## OCTEON 10 platform initial family members

<table>
<thead>
<tr>
<th>Metric</th>
<th>CN103XX</th>
<th>CN106XX</th>
<th>CN106XXS</th>
<th>DPU400</th>
</tr>
</thead>
<tbody>
<tr>
<td>N2 Cores</td>
<td>Up to 8</td>
<td>Up to 24</td>
<td>Up to 24</td>
<td>Up to 36</td>
</tr>
<tr>
<td>Max Frequency</td>
<td>2.5GHz</td>
<td>2.5GHz</td>
<td>2.5GHz</td>
<td>2.5GHz</td>
</tr>
<tr>
<td>SPECint (2006)</td>
<td>&gt;275</td>
<td>&gt;800</td>
<td>&gt;800</td>
<td>&gt;1200</td>
</tr>
<tr>
<td>Cache (L2, L3)</td>
<td>8MB, 16MB</td>
<td>24MB, 48MB</td>
<td>24MB, 48MB</td>
<td>36MB, 72MB</td>
</tr>
<tr>
<td>DDR5 Controllers</td>
<td>2 at 4800MT/s</td>
<td>6 at 5200MT/s</td>
<td>6 at 5200MT/s</td>
<td>12 at 5200MT/s</td>
</tr>
<tr>
<td>Crypto</td>
<td>Supported</td>
<td>Supported</td>
<td>Supported</td>
<td>Supported</td>
</tr>
<tr>
<td>Ethernet</td>
<td>4x50G/25G/10G + 2x10G or 16x1G</td>
<td>4 x50G or 2x10/1G</td>
<td>16 x50G</td>
<td>Up to 400G</td>
</tr>
<tr>
<td>PCIe 5.0 controllers</td>
<td>Up to 6</td>
<td>Up to 6</td>
<td>Up to 4</td>
<td>Up to 8</td>
</tr>
<tr>
<td>Typical power</td>
<td>10-25W</td>
<td>40W</td>
<td>50W</td>
<td>60W</td>
</tr>
</tbody>
</table>

*Sampling in 2H21*
Essential technology, done right™
Thank You