AMD 400G Adaptive SmartNIC SoC
Technology preview

Jaideep Dastidar
Co-authors: David Riddoch, Jason Moore, Steve Pope, Jim Wesselkamper

Hot Chips 2022
**Overview**
- Motivating factors behind SmartNICs
- AMD 400G Adaptive SmartNIC SoC Objectives

**Solution – AMD 400G Adaptive SmartNIC SoC**
- Offload Acceleration Blocks for Host, Network, and Embedded Processing
- Adaptive Extensions
- Pervasive Security and Confidential Computing

**Use Case Examples**
- OVS Offload
- VirtIO Net Offload
- Host-Attached Adaptive Network Offload Accelerator
- Network Attached Adaptive NVMe Storage Endpoint

**Summary**

**Q&A**
Overview
Motivations for Adaptable, Intelligent Infrastructure

New Trends Expose Current Infrastructure Challenges

Intelligent Infrastructure for Next Gen Applications

Next Generation Bandwidth
400Gb Data Processing

Infrastructure Acceleration
Maximize CPU revenue potential

High Security
Multi-tenant Isolation and Protection
SmartNIC Solution Space

- Proprietary
- Global Hyperscalers
- Evolving Standards
- Telco/CDN
- Customization
- Other CSPs
- Enterprise
- Zer-touch

- SR-IOV
- OVS
- HW Virtio.net / blk
- IPsec
- Overlays
- SR-IOV
- DPDK
- kTLS
- vDPA
- Orchestration
We are at an Architectural Inflection Point

### Tug-of-War between

**ASIC-like Fixed Function Logic**
Frequent vs occasional functions

**Embedded Processor Cores**
Workload flexibility vs processing offload

**FPGAs or Programmable Logic**
Adapting new and existing functions
Balanced Architecture
• Architect an adaptive SoC for the SmartNIC domain
• Strike the right balance between the heterogeneous elements
• Present a unified software view of the heterogeneous element SoC

Performance and Adaptability
• Latest Interfaces/Speeds: 112G Ethernet, PCIe Gen5, CXL 2.0, LPDDR5/DDR5
• Adaptive Interface APIs for evolving workloads and customization

Security
• Latest security standards: PCIe, OCP and NIST
• Symmetric and asymmetric crypto offload for data encryption and key exchange
• Physical and logical isolation for confidential compute
AMD 400G Adaptive SmartNIC SoC

Technology preview
Solution

ASIC  ↔  Software  ↔  Prog. Logic

Co-design

ASIC Logic

Common Memory & NoC

Crypto Offload

DMA Offload

Network Dataplane Offload

ASIC ↔ PL Adapters

ASID Logic

Programmable Logic

SW ↔ PL Adapters

SW ↔ ASIC Adapters

Telemetry

Network Control Plane

Custom Coherent Accelerators

Crypto Offload

Custom Dataplane Offload

Custom Header Extensions

Hot Add / Remove New Functions
Tightly Coupled SmartNIC Blocks

- **Host Subsystem** (CPM5N)
- **Processing Subsystem** (PSX)
- **Programmable Logic**
- **Network Subsystem** (HNICX)
- **Memory Subsystem**
CPM5N: Host Subsystem Overview

PCIe Gen 5 or CXL 2.0 host connection

Composable DMA (CDx)

Host DTE (Data Transform Engine)
HNICX: Networking Subsystem Overview

Hardened NIC (HNIC)

Network DTE
(Data Transform Engine)

MAC
Memory Subsystem Overview

8x LPDDR5 / DDR5 Controllers
- 32-bit DDR5-5600 with ECC or
- 16-bit, 32-bit, or dual 160-bit LPDDR-6400 with ECC
- Inline Memory Encryption
  - AES-GCM / AES-XTS with DPA countermeasures

DDR Controller Bypass Mode for SCM
- PHY also supports other standards such as MIPI
- Custom PL Controllers for Storage Class Memory
## Processing Subsystem (PSX) Overview

<table>
<thead>
<tr>
<th>LPD</th>
<th>FPD</th>
</tr>
</thead>
</table>
| **Cortex-R52**  
| 32KB I/D $  
| 128KB TCM | **Cortex-A78AE**  
| 64KB I$/D$  
| 512KB L2$ |
| LS/Split | LS/Split |

**4 real-time cores arranged as Two 2-core Clusters w/ Tightly-Coupled Memory**

| **Cortex-R52**  
| 32KB I/D $  
| 128KB TCM | **Cortex-A78AE**  
| 64KB I$/D$  
| 512KB L2$ |
| LS/Split | LS/Split |

**High Performance Domain**

16 Application cores arranged as Four 4-core Clusters w/Crypto Extensions

### Interconnect

- 1MB OCM
- GbE (2)
- USB2 (2)
- SPI (2)
- MIO
- PSM
- PS Mgmt
- DMA
- CANFD (2)
- UART (2)
- I2C+I3C (2)
- Timers
- Debug
- PS CoreSight

### Coherency Mesh Network

- LLC 16MB

### Core Sight

- System MMU
- Interconnect
- GIC
- PKI

### Offload Acceleration

Public Key Infrastructure TLS 1.3
Tight Coupling between CPM5N ↔ PSX

256Gbps Host MMIO to Embedded Cores Data Flow w/ Cache Injection via PCIe/CXL.io Endpoint

256Gbps Embedded Cores to Host Data/Doorbells

Host Subsystem
(CPM5N)

Processing Subsystem
(PSX)
Tight Coupling between Prog. Logic $\leftrightarrow$ PSX

512Gbps Near-PL Fabric I/O Coherent access w/ Cache Injection and Maintenance
256Gbps Embedded Core MMIO access to Near-PL Memory

Processing Subsystem (PSX)

256Gbps Cache Coherent Near-PL Fabric access including Cache Coherent ATOMICS

Programmable Logic
Security and Confidential Compute Across Domain-specific Accelerators

Shu: First, I want to say I really liked what you did on slide 16, 17, and 18. However, we've had some feedback from our legal department. Unfortunately, they have us they do not want us to use locks or shields. They indicated that some imagery that included castles have worked. The idea portrayed along with the castle imagery is "layered security features". A castle has many layers of protections around it. The castle seems to imply less that it can’t be broken into, but there are several layers of protection around it, including a moat, several walls, etc. To be honest… I’m not sure "castle" images go with this presentation… but maybe you have something abstract enough that it doesn’t look like it is from 500 years ago.
Pervasive Security
Layers of Security

CHECK
Secure Boot, Key Exchange and Attestation
AMD keys, SmartNIC owner keys, and SmartNIC tenant keys

PROTECT
Memory and Peripheral Protection Units
SoC-wide hardware firewalls for tenants and accelerators

SHIELD
Secure Data in Flight, at Rest, in Use
Encryption for data, host/network interfaces, and DRAM
Secure monitors for external, internal, physical attacks
Crypto engines with DPA countermeasure protection
Use Case Examples
Use-cases

Network Attached

Host Attached

Host Attached

Storage Attached
Full OVS Offload

- Packet switching and packet modifications
- Overlay encapsulation and decapsulation
- Firewall offload including connection tracking / NAT
- Packet monitoring
- Uses BCAM / TCAM to match packet flows
- These flow tables stored in DRAM (with internal caching)
- OVS control plane remains in software running on the host
VirtIO-net Support

Device Natively Supports VirtIO Queues

Two Deployment Models
- Bare Metal VirtIO
- vDPA
  (Virtual Data Path Acceleration)
PCIe Endpoint 400G SmartNIC w/ with Host Storage and Network Offload

- Standard Ethernet Packet Flow
- Custom Header, Encapsulation, or Encryption Flow
- Host Connection with Advanced Switch Hot Plug
PCle Endpoint 200G SmartNIC w/ 2 Gen5 x4 NVMe SSDs

- Host endpoint connection NVMe virtualization
- SmartNIC Root-Port connections for dual-NVMe SSDs w/ AES-XTS offload acceleration
- 2x 100G or 4x 50G network connection
PCle Root Complex

400G NVMe Storage Node

- Network-attached storage appliance
- SmartNIC Root-Port connections for quad-NVMe SSDs w/ AES-XTS offload acceleration
- 2x 200G or 4x 100G remote storage connections
Summary
## Performance & Resources

### Manufacturing Technology

<table>
<thead>
<tr>
<th>Process Technology</th>
<th>TSMC 7nm FF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Host Subsystem (CPM5N)</td>
<td>Programmable NoC</td>
</tr>
<tr>
<td>Network Subsystem (HNICX)</td>
<td>Programmable Logic</td>
</tr>
<tr>
<td>Processing Subsystem (PSN)</td>
<td>Memory Subsystem</td>
</tr>
</tbody>
</table>

### Architectural Resources

<table>
<thead>
<tr>
<th>Subsystem</th>
<th>Resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Host subsystem</td>
<td>PCIe Gen5 x16, PCIe CMA/DDE/IDE, CXL v2.0</td>
</tr>
<tr>
<td>Network subsystem</td>
<td>2x 200G Ethernet, Virtual switch offload, Stateless offloads</td>
</tr>
<tr>
<td>Application Processing Subsystem</td>
<td>16x A78-AE, 24MB Shared Cache</td>
</tr>
<tr>
<td>Real-time Processing Subsystem</td>
<td>4x ARM R52</td>
</tr>
<tr>
<td>PKI (TLS1.3 Offload Engine)</td>
<td>Yes</td>
</tr>
<tr>
<td>DDR subsystem</td>
<td>8x 32b LPDDR5/DDR5, ECC, Inline encryption, 1.638Tb/s LPDDR5 BW</td>
</tr>
</tbody>
</table>

### SmartNIC SoC Performance†

<table>
<thead>
<tr>
<th>Performance</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Programmable logic packet rate¹</td>
<td>400Mpps Ingress + 400Mpps Egress</td>
</tr>
<tr>
<td>Host ⇔ Network packet rate²</td>
<td>400Mpps RX + 400Mpps TX</td>
</tr>
<tr>
<td>Full Virtio.NET offload BW³</td>
<td>400Gbps RX + 400Gbps TX</td>
</tr>
<tr>
<td>AES-XTS offload BW⁴</td>
<td>800Gb/s</td>
</tr>
<tr>
<td>AES-GCM offload BW⁵</td>
<td>800Gb/s</td>
</tr>
</tbody>
</table>

† See Endnotes
Thank You

To the worldwide AMD AECG team and all the other AMD teams who made the AMD 400G Adaptive SmartNIC possible

Their creativity, ingenuity, and unwavering dedication to the project has made for a unique Adaptive SmartNIC DSA
Endnotes

† SmartNIC SoC performance claims are based on simulation results of pre-silicon models with AMD traffic generators used to emulate external port traffic behavior as of August 23, 2022. Actual performance of production silicon may vary.

1. Interface packet rate between Programmable logic and the Network subsystem
2. Sustained TX and RX packet rate for 64B packets
3. Sustained TX and RX BW with 1500B payload per packet
4. Sustained BW for 1KB packets; Encrypt-only, Decrypt-only, or 400Gb/s Encrypt + 400Gb/s Decrypt with Network Interface configured to 2x200G Ethernet ports
5. Sustained BW for 1KB packets; 400Gb/s Encrypt + 400Gb/s Decrypt with Network Interface configured to 2x200G Ethernet ports
6. #4 and #5 were also measured concurrently enabled
Disclaimer and Attribution

The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions and typographical errors.

The information contained herein is subject to change and may be rendered inaccurate for many reasons, including but not limited to product and roadmap changes, component and motherboard version changes, new model and/or product releases, product differences between differing manufacturers, software changes, BIOS flashes, firmware upgrades, or the like. AMD assumes no obligation to update or otherwise correct or revise this information. However, AMD reserves the right to revise this information and to make changes from time to time to the content hereof without obligation of AMD to notify any person of such revisions or changes.

AMD MAKES NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE CONTENTS HEREOF AND ASSUMES NO RESPONSIBILITY FOR ANY INACCURACIES, ERRORS OR OMISSIONS THAT MAY APPEAR IN THIS INFORMATION.

AMD SPECIFICALLY DISCLAIMS ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR ANY PARTICULAR PURPOSE. IN NO EVENT WILL AMD BE LIABLE TO ANY PERSON FOR ANY DIRECT, INDIRECT, SPECIAL OR OTHER CONSEQUENTIAL DAMAGES ARISING FROM THE USE OF ANY INFORMATION CONTAINED HEREIN, EVEN IF AMD IS EXPRESSLY ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

ATTRIBUTION

© 2022 Advanced Micro Devices, Inc. All rights reserved. AMD, the AMD Arrow logo, the Xilinx logo, Versal, and combinations thereof are trademarks of Advanced Micro Devices, Inc. in the United States and/or other jurisdictions.
Thank You for Participating