Cortex-A72: Current State-of-the-Art Processor

Aniket M. Saha
Senior Product Manager, ARM
Cortex-A72: State of the Art Processor

- **Compelling single-threaded performance**
  - Large performance increase across all workloads including integer, memory-intensive, crypto, floating point, etc.
  - Baseline microarchitecture similar to Cortex-A57

- **Significant advancements in power efficiency**
  - Re-optimized every logical block from Cortex-A57
  - Power reduction enables sustained operation at Fmax
  - Area reduction lowers costs and static power

- **Feature support for enterprise and mobile SoCs**
Cortex-A72: ARM's Highest Performance Processor

Increase in **sustained performance** within smartphone power budget

<table>
<thead>
<tr>
<th>Year</th>
<th>Cortex</th>
<th>Process</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014</td>
<td>Cortex-A15</td>
<td>28nm</td>
<td>1.6 GHz</td>
</tr>
<tr>
<td>2015</td>
<td>Cortex-A57</td>
<td>20nm</td>
<td>2.0 GHz</td>
</tr>
<tr>
<td></td>
<td>Cortex-A57</td>
<td>14/16nm</td>
<td>2.3 GHz</td>
</tr>
<tr>
<td></td>
<td>Cortex-A72</td>
<td>14/16nm</td>
<td>2.5 GHz</td>
</tr>
</tbody>
</table>

3.5x

Premium
Cortex-A72: Next-Generation Performance

Workloads include: SPECint06, SPECfp06, Stream, LMbench, Geekbench, Antutu, Minebench, AES/SHA/CRC kernels, and other targeted kernels
Next generation solutions using ARM Cortex-A72
Enabling Scalable Portfolio of Solutions

Cortex-A7
Cortex-A53
Cost-Efficient Power-Optimized
CCI-400
CCI-500
CCN-502

Cortex-A53
Cortex-A57
Mid-range Performance
CCI-500
CCI-502
CCN-504

Cortex-A53
Cortex-A57
Cortex-A72
High Performance Networking and Server
CCI-508
CCN-512
Extensible Architecture for Heterogeneous Multi-core Solutions

- Up to 4 cores per cluster
- Heterogeneous processors – CPU, GPU, DSP and accelerators
- Virtualized Interrupts
- Up to 12 coherent clusters
- Integrated L3 cache
- Up to Quad channel DDR3/4 x72

Peripheral address space

Cortex A72
Cortex A73
Cortex A72
Cortex A73
Cortex A72
Cortex A73
Cortex A72
Cortex A73

Memory Controller DMC-520
x72 DDR4-3200
x72 DDR4-3200
x72 DDR4-3200
x72 DDR4-3200

PCIe
SATA
USB
DPI
Crypto
DSP
NIC-400

Up to 24 I/O coherent interfaces for accelerators and I/O

Peripheral address space

A-32MB L3 cache

Snoop Filter

Network Interconnect NIC-400

Flash
SRAM
GPIO
PCIe

Network Interconnect NIC-400

I/O Virtualisation CoreLink MMU-500

ARM®

Cortex A72
Cortex A53
Cortex A72
Cortex A53
Cortex A72
Cortex A53
Cortex A72
Cortex A53

GIC-500

I/O Virtualisation CoreLink MMU-500

I-32MB L3 cache

Integrated L3 cache

Up to 12 coherent clusters

Up to 4 cores per cluster

Virtualized Interrupts

Heterogeneous processors – CPU, GPU, DSP and accelerators

ARM®

Cortex A72
Cortex A53
Cortex A72
Cortex A53
Cortex A72
Cortex A53
Cortex A72
Cortex A53

GIC-500

I/O Virtualisation CoreLink MMU-500

I-32MB L3 cache

Integrated L3 cache

Up to 4 cores per cluster

Virtualized Interrupts

Heterogeneous processors – CPU, GPU, DSP and accelerators

ARM®
### Efficient Interconnect for Compelling Scalable Solutions

<table>
<thead>
<tr>
<th>Cost-efficient</th>
<th>Mid-range</th>
<th>High-end</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level-3 Cache Size</td>
<td>32MB</td>
<td></td>
</tr>
<tr>
<td>DDR Bandwidth</td>
<td>100 GB/s</td>
<td></td>
</tr>
<tr>
<td>Coherent ports</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>On-chip bandwidth</td>
<td>1.8 Tb/s</td>
<td></td>
</tr>
</tbody>
</table>

**Approximate core count**

- **CCI-400**
- **CCI-500**
- **CCN-502**
- **CCN-504**
- **CCN-508**
- **CCN-512**

**System Performance**

- **AMBA 5 CHI**
- **AMBA 4 ACE**
Enterprise Compute Requirements

Control Plane Processing
- Fast Event Processing
- Complex signalling
- Trend: Evolving Software

Need: Efficient, High Compute Performance

MAC Scheduling
- Real Time, Latency Driven
- Multiple core processing
- Trend: More Complexity (LTE-A, 5G)

Need: High Compute, Low Latency Performance

High Bandwidth, Low Latency Interconnect
Wide Range of Implementations from Few to Many Coherent Devices

Data Plane Processing
- Throughput driven, IO intensive
- Deterministic performance
- Trend: Higher packet rates

Need: Small Cores at Maximum Efficiency

Specialised Processing
- L1, Content Delivery, Security
- Diverse requirements
- Trend: Advanced modulation schemes

Need: DSPs, Accelerators
Cortex-A72: Compelling performance and throughput

ARM Cortex-A57, Cortex-A72 deliver:

- Competitive performance per thread
- Similar overall performance throughput
  - At Much Lower Power

Comparison for equivalent number of threads
- Platforms used:
  - Xeon-E5 2660 10C20T platform (measured)
  - Xeon-E5 2650 10C20T platform (measured)
  - Gcc compiler v4.9 with -o3 flag

- Estimated result on example 20C ARM Cortex platforms with CCN-508, 28MB total L2+L3 cache
  - per-core measurements on RTL with relevant memory system
  - Gcc compiler v4.9 with -o3 flag
  - Scaled to 20T based on modelled and empirical results
  - Power estimated in 16nm based on ARM internal implementations for entire CPU+interconnect complex

<table>
<thead>
<tr>
<th>Workload</th>
<th>Relative Performance (Spec2K6 rate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xeon-E5 2650 V3</td>
<td>2.3 GHz (10 cores 20 threads)</td>
</tr>
<tr>
<td>Cortex-A57</td>
<td>2.5 GHz (20 cores 20 threads)</td>
</tr>
<tr>
<td>Cortex-A72</td>
<td>2.7 GHz (20 cores 20 threads)</td>
</tr>
<tr>
<td>Xeon-E5 2660 V3</td>
<td>2.6 GHz (10 cores 20 threads)</td>
</tr>
</tbody>
</table>
Maximizing Throughput Density: per mm², per Watt

ARM Solution Benefits:

- Less than 1/3rd the power for equivalent performance
- Allows more specialized computing or significantly greater thread density in the same power budget

Comparison for equivalent number of threads

Platforms used:
- Xeon-E5 2660 10C20T platform (measured)
- Xeon-E5 2650 10C20T platform (measured)
- Gcc compiler v4.9 with –o3 flag

Estimated result on example 20C ARM Cortex platforms with CCN-508, 28MB total L2+L3 cache
- per-core measurements on RTL with relevant memory system
- Gcc compiler v4.9 with –o3 flag
- Scaled to 20T based on modelled and empirical results
- Power estimated in 16nm based on ARM internal implementations for entire CPU+ interconnect
**Cortex-A72: Ideal for dense compute environments**

<table>
<thead>
<tr>
<th>Description</th>
<th>Diagram</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single Cortex-A72 core</td>
<td><img src="image1" alt="Diagram" /></td>
</tr>
<tr>
<td>~1.15mm(^2)</td>
<td>Cortex-A72 is (&lt;20%) size</td>
</tr>
<tr>
<td>Single Broadwell CPU + 256K(^1) L2</td>
<td><img src="image2" alt="Diagram" /></td>
</tr>
<tr>
<td>~8mm(^2)</td>
<td>A quad core Cortex-A72 with 8x L2 cache RAM is the same size</td>
</tr>
</tbody>
</table>

\(^1\)Source: Estimated from die-shot image provided by Intel at IDF 2014.  
\(^2\)Source: ARM trial implementations on TSMC 16FF+, using ARM Artisan libraries
Cortex-A72: Highest Performance ARM Cortex Processor

- Compelling performance and efficiency
- Enterprise class scalable solutions
- Enterprise ready feature set and ecosystem