Integrating CPU and GPU, The ARM Methodology

Steve Steele, Senior Product Manager
Media Processing Division, ARM
Mobile: A Growing Part of Everyday Life

- 60% Facebook users access from mobile
- 36% Chinese Android users spend 2+ hours a day on mobile apps
- 192% Increase in mobile share of web traffic in Asia (2010-2012)
- Nearly 1/3 of UK page views are from mobiles and tablets
- 200 Million Active Twitter users, 60% via Mobile
- >48% of web traffic is mobile in India
- 25% Of YouTube views come from mobile

The Architecture for the Digital World®
Higher-End Features for New Use-Cases

Video editing and effects creation:
High-performance processor; video and graphics

Natural language and handwriting recognition:
High-efficiency processing; GPU compute; security

Facial recognition:
High-efficiency processing
GPU compute
Security

Bring Your Own Device, Mobile Payments, and Digital Identity
Hardware-based security and Virtualization

The Architecture for the Digital World®
Tuning Solutions for each Market

Mass market smartphones
- Cortex-A7 Mali-T628
- Cortex-A7 Mali-T624

High-end smartphones
- Cortex-A7 & Cortex-A57
- Cortex-A15 & Next gen Mali

Embedded
- Cortex-M Series
- Cortex-R Series

Servers
- Cortex-A57

Networking
- Cortex-A15

Mobile computing & DTV
- Next gen Mali

The Architecture for the Digital World®
Market Requirements Evolve Over Time

**Market trends**
- Ever increasing user demands
- Adoption of OpenGL ES 3.0 and GPU Compute becomes mainstream
- Wide number of UI & Gaming use cases
- Adoption of OpenGL ES 3.0 in premium devices
- Diversifying OS support – Android still dominant
- GPU Compute enabled features start to enter market

**Mali GPU features**
- System compute libraries across NEON and GPU
- Extended coherency & power management
- Extensive ecosystem reduce TTM
- Scalable Drivers
- System level efficiencies (ASTC, AFBC)
- Scalable solutions leveraging one Ecosystem
- Proven CPU & GPU system approach
- System wide tools (DS-5)
- Area and power optimized POP
Solving the Challenges for System Design

- Increasing resolution
- Increasing complexity
- Natural UIs (gesture control)
- µArchitectural improvements
- Software efficiencies
- System optimizations

A Balance that must keep within the Thermal Limit of the device
The Chip is the System

- ARM delivers technology to drive efficient system-on-chip solutions:
  - **Software** increasing system efficiency with optimized software solutions
  - Diverse components, including CPU and GPU processors designed for specific tasks
  - **Interconnect System IP** delivering coherency and the quality of service required for lowest memory bandwidth
  - **Physical IP** for a highly optimized processor implementation
ARM Technology

- Advanced consumer products are incorporating more and more ARM technology – from processor and multimedia IP to software

Processor IP – Design of the brain of the chip
Graphics IP – Design of the graphics muscle of the chip
Physical IP – Design of the building blocks of the chip
Software development tools
One Product Does Not Fit Every Need

- Advanced graphics for Consumer Entertainment & Mobile
  - Higher screen resolutions
  - Richer, more complex user interfaces, applications & games
- Graphics and GPU Compute for Mobile Computing
  - Enabling next generation use cases
  - Combine Cortex and Mali processors into unified compute sub-system
- Roadmap aligned with market to give uncompromised choice – balancing:
  - Die area (cost and yield)
  - GPU Compute capability
  - Software requirements
  - High performance
  - Energy efficiency
Best for Graphics Only

- The optimum balance of graphics performance, power and cost
- Multicore delivers performance scalability over many form factors
- Suitable for the fastest growing entry-level and mid-range smartphone markets
- A common software platform reduces system cost and TTM
- Mali-450 MP will drive volume in next generation products in CE and mobile markets

Performance

Mali-450 MP
- 2x Mali-400 MP performance
- Scalable to 8 cores
- Leading OpenGL ES 2.0 performance

Mali-400 MP
- First OpenGL ES 2.0 multi-core GPU
- Scalable to 4 cores
- Leading energy-efficiency

Mali-300
- Entry-level OpenGL® ES 2.0 GPU

The Architecture for the Digital World®
Best for Graphics and GPU Compute

- Designed for GPU Compute
  - Uncompromised support for OS / API choice
  - Full Profile, 64-bit Compute, Double-precision IEEE-754-2008

- Closer CPU-GPU links
  - Efficient use of all device resources
  - Maximize performance and battery life
  - Coherent memory links
  - Right task in the right place

- Protecting partner investments
  - Common software platform reduces costs and TTM
  - Multicore delivers performance scalability over multiple form factors

- Advanced products in market early
  - Mali-T604 silicon shipping now in consumer products

Performance

- Mali-T628
  - High end solution
  - Max performance
  - Optimized for high-end

- Mali-T624
  - Doubles performance
  - Scalable to 4 cores

- Mali-T622
  - 50% energy efficiency improvement
  - ASTC support

- Mali-T604
  - First Midgard architecture product
  - OpenGL ES 3.0 support
  - Scalable to 4 cores

Date of production chips:

- 2011 Mali-T604
- 2012 Mali-T622
- 2013 Mali-T624
- 2014 Mali-T628

Product is based on a published Khronos Specification, and is expected to pass the Khronos Conformance Testing Process. Current conformance status can be found at www.khronos.org/conformance
Comprehensive GPU Compute Support

- ARM’s best-in-class CPU know-how combined with expertise in graphics technology enabling complex use-cases
  - Computational photography: Panorama stitching
  - Image recognition: Face, smile, landmark, context
  - Image improvement, stabilization, editing, filtering

- By moving GPU Compute tasks onto the GPU will enable lower power consumption and faster response over being solely run on the CPU
Mali GPU Compute: No FUD... Facts

- Passed Khronos Conformance
  - OpenCL™ 1.1 Full Profile on Linux and Android™
- Proven in Silicon
  - Samsung Exynos 5 Dual, implements Full Profile
  - OpenCL and Renderscript DDK available now
- Mali-T604 shipping in real products
  - Google Chromebook
  - Google Nexus 10
  - InSignal Arndale Community Board
- API exposed for developers
  - OpenCL on Linux for Arndale platform
  - Renderscript computation on Android for Nexus 10
Compute Use Case Example

- ARM internal experience
- OpenCL 1.1 FP accelerated world
  - Interactive items and lights
  - Bullet physics broad-phase fully OpenCL accelerated on GPU
- Performance boost
  - GPU Kernel speedup >10x
  - But system speedup is less
- ARM integration goal
  - *Take the system cost out!*
Integration: Coherency

- SoCs are heterogeneous systems
- But sharing data can still be costly
  - Cache flushes, locks, syncs reduces the heterogeneous benefit
- HW coherency makes sharing data cheap and automatic
- ARM is in leading position with full technology coverage
  - Cortex™ CPUs
  - Mali GPUs
  - CoreLink™ system IP
  - AMBA™ bus protocols
Integration: Address Space Alignment

- The 32-bit address space is running out, even in mobile
- Midgard architecture built for full 64-bit addresses
- Embedded distributed Mali MMU for VA to PA/IPA translation
  - Mali-T604: 48-bit VA and 40-bit PA/IPA
  - Uses ARMv7 LPAE page table format, just like Cortex-A15 & Cortex-A7
- Multiple simultaneous address spaces supported
  - Mali GPUs run many threads in parallel
  - Independent processes may execute on GPU simultaneously
  - Seamless process transitions ensures maximum utilization/efficiency

64-bit Pointers

48-bit VA

40-bit IPA/PA
ARM System Scalability

Introducing CCI-400 Cache Coherent Interconnect

- Processor to Processor Coherency and I/O coherency
- Memory and synchronization barriers
- Virtualization support with distributed virtual memory signaling

Quad Cortex-A15 MPCore
- A15, A15, A15, A15
- Processor Coherency (SCU)
  - Up to 4MB L2 cache

Quad Cortex-A7 MPCore
- A7, A7, A7, A7
- Processor Coherency (SCU)
  - Up to 4MB L2 cache

Mali-T624
- GPU Core, GPU Core, GPU Core, GPU Core
- Mali L2 Cache

128-bit AMBA 4

CoreLink CCI-400 Cache Coherent Interconnect
Leadership in Lowering System Power

- The GPU has a major impact on the SoC architecture
  - Area, memory bandwidth, energy and implementation
- ARM focuses on system-wide power efficiency not just the individual IP components
- Energy saving features in the Mali-T62x system include:
  - 50% GPU performance efficiency increase or less energy/frame in same area
  - Transaction Elimination 50% of the total GPU bandwidth
  - ASTC Texture Compression 90% texture bandwidth reduction while preserving quality
- ARM POP™ IP for Cortex CPUs and Mali GPUs
  - Up to 27% higher frequency
  - 24% lower area
  - 19% lower power
Delivering Full-featured, Mid-Range Devices

- Balancing your budget (cost and power)
  - Screen size
  - Display resolution
  - Battery capacity
  - Camera quality
  - Storage size and RAM
  - Look and feel
  - Modem technology
  - SoC

- Minimize development costs and time to market

- Delivering premium features within the budget → optimized SoCs

Premium features, Highest specifications

Premium features, Range of specifications, Diversity of devices

The Architecture for the Digital World®
ARM’s Newest Optimized IP Solutions

System Solution for Mid-range Market:

- ARM Cortex-A12 CPU
- ARM Mali-T622 GPU
- ARM Mali-V500 video
- ARM POP IP for CPU and GPU at 28nm

Mali
Visual Computing by ARM

Cortex
Low-Power Leadership from ARM

Artisan
Advanced Physical IP by ARM

The Architecture for the Digital World® ARM
ARM Cortex-A12 CPU

- 40% performance uplift over Cortex-A9
- Same best-in-class energy efficiency
  - The most area- and cost-efficient solution
- Premium mobile features
  - big.LITTLE™ processing enabled
  - Greater than 4GB addressable memory
  - Security with Virtualisation and TrustZone®
ARM Mali-T622 - Entry Level GPU Compute

- Innovative GPU architecture
  - Designed for performance and flexibility
- Leading graphics performance
  - Support for OpenGL ES 3.0 and ASTC for the richest user-experience
- Architected for GPU computing - the most efficient GPU Compute solution in the market
  - Entry level OpenCL 1.1 Full Profile solution
  - Renderscript Compute for mid-range smartphones
- Energy and bandwidth optimizations
  - 50% energy efficiency improvements
  - Increased efficiency for compute functions
  - Transaction elimination and Hierarchical tiling
- A system approach
  - ARM CPU + GPU + Memory + Interconnect
  - ARM POP IP for Mali and ARM DS-5 support

![Computer vision](image1)

![Video stabilization](image2)

The Architecture for the Digital World®
Video Trends

- **HD Video Everywhere**
  - 2/3 of world's mobile data traffic will be video by 2017\(^1\)
  - 66% of mobile data traffic will be video by 2017\(^1\)
  - 25% of global YouTube views come from mobile devices\(^2\)

- **Video and Graphics both consume significant amounts of memory bandwidth**
  - Lower bandwidth = lower power consumption

- **High Quality Required Across All Screens**
  - High picture quality with no dropped frames
  - WiFi display connecting mobile and TV

- **Major Movie studios demanding HW protection of assets - from download to display**

---

ARM Mali-V500 – Mass Market HD Video

- System perspective to power reduction
  - 50% lower memory bandwidth with ARM Frame Buffer Compression (AFBC) in Video+Display
  - Small area Video for lowest cost and power
  - 1080p60 encode/decode to 4K120

- Optimized for real embedded systems
  - Latency tolerant design – no dropped frames
  - High quality IP from ARM

- Secure video path
  - Optimized for TrustZone secure media use cases
ARM POP IP for Cortex-A12 and Mali-T622

- Best PPA and Faster Time to Market for ARM Cortex CPUs and Mali GPUs
- Implementation Flexibility
  - Performance, Power, Area

<table>
<thead>
<tr>
<th>Foundry</th>
<th>Process</th>
<th>Cortex POP IP</th>
<th>Mali POP IP</th>
</tr>
</thead>
<tbody>
<tr>
<td>GlobalFoundries</td>
<td>GF28-SLP</td>
<td>Cortex-A12</td>
<td>-</td>
</tr>
<tr>
<td>TSMC</td>
<td>28HPM</td>
<td>Cortex-A12</td>
<td>Mali-T622</td>
</tr>
</tbody>
</table>

The Architecture for the Digital World®
Dynamic and Growing Mali Ecosystem

**USE CASES & SEGMENTS**

- USER INTERFACE
- BROWSER
- GAMING
- COMPUTATIONAL PHOTOGRAPHY
- COMPUTER VISION
- MOBILE & TABLET
- HOME ENTERTAINMENT

**DEVELOPER RESOURCES**

- SDKS
- EMULATORS
- FAST MODELS
- TEXTURE TOOLS
- SHADER & KERNEL TOOLS
- GRAPHICS & COMPUTE
- DEBUG
- PERFORMANCE ANALYSIS

**DEMOS, EVENTS & DEVREL**

- INTERNAL DEMOS
- COMMISSIONED DEMOS
- PARTNER DEMOS
- EVENTS PLANNING & SUPPORT
- H/W & BSPS
- LOGISTICS & ADMIN
- DEVELOPER SUPPORT

**DEVELOPER EDUCATION**

- PRESENTATIONS
- WORKSHOPS
- TUTORIALS
- SAMPLE CODE
- WEBSITE
- COMPETITIONS & HACKATHONS
- DEVELOPER GUIDES
The Mali Ecosystem – Flexible Partner Choice

- ARM GPU roadmap enables partner choice and flexibility on software
- Partnerships in place to enable a leadership position
System-Level Developer Benefits

- Developers need to see the workload across the whole CPU + GPU system
  - Because applications do not run in isolation on a single processor
  - Complex interaction of components executing on both CPU and GPU
  - DS-5™ toolchain provides a system-level view across CPUs and GPUs

- Speed-up development time
- Maximize performance
- Reduce power consumption
ARM DS-5 Streamline Performance Analyzer

- API Events
- CPU Activity
- GPU Activity
- Filmstrip
- S/W Counters
- H/W Counters
- Heatmap
ARM® Mali™ GPU Momentum

Over 50% of Android tablets Mali-based
Over 20% of Android smartphones Mali based
Over 70% of smartTVs Mali-based
Over 150M Mali GPUs shipped in 2012
Licensing grows by 30% to 75 licenses
Shipment growth continues in line with 2012 growth
Samsung launched Mali-T604 based F8000 at CES
Over 70% of smartTVs Mali-based

Mali GPU Shipments Outpace Industry Growth
Mali is Gaining Market Share

Cortex®-A9 and Mali-400 found in a wide range of DTV & STB

The Architecture for the Digital World®
Summary

- Getting the maximum efficiency from modern SoCs is highly complex
  - Interactions between many sub-systems to consider and optimize
  - Requires new innovations and a technology focus
- ARM Cortex CPU / coherent Mali GPU / big.LITTLE enable highest performance and scalability from mobile through to console class gaming
- ARM continues to drive the development for better system integration
  - Cortex™ CPUs, Mali™ GPUs and CoreLink™ fabric leading the way
  - Development Studio 5 (DS-5™) leading edge support for ARM technologies
- The ARM Methodology - it is the winning formula
謝謝您
Thank You