DELIVERING A NEW LEVEL OF VISUAL PERFORMANCE IN AN SOC
AMD “RAVEN RIDGE” APU

Dan Bouvier, Jim Gibney, Alex Branover, Sonu Arora

Presented by:
Dan Bouvier
Corporate VP, Client Products Chief Architect
RAISING THE BAR FOR THE APU VISUAL EXPERIENCE

FIRST
“Zen”-based APU

HIGH-PERFORMANCE
On-die “Vega”-based graphics

LONG BATTERY LIFE
Premium form factors

Up to 200% MORE CPU PERFORMANCE

MOBILE APU GENERATIONAL PERFORMANCE GAINS

Up to 128% MORE GPU PERFORMANCE

Up to 58% LESS POWER

CPU Performance | GPU Performance | Power

Scaled GPU and CPU up to reach target frame rate
Managed power delivery and thermal dissipation
Improved memory bandwidth efficiency
Upgraded display experience
Increased package performance density

* See footnotes for details.
"RAVEN RIDGE" APU

AMD "ZEN" x86 CPU CORES

CPU 0
(4 CORE | 8 THREAD)

CPU 2
4MB L3 Cache

CPU 1

CPU 3

X64 DDR4

Infinity Fabric

System Management Unit

AMD GFX+
(11 COMPUTE UNITS)

1MB L2 Cache

CU

Display Controller
Next

Platform Security Processor

Multimedia Engines

Video Codec Next

Audio ACP

Sensor Fusion Hub

UPGRADED DISPLAY ENGINE

AMD "VEGA" GPU

FULL SYSTEM CONNECTIVITY

ACCELERATED MULTIMEDIA EXPERIENCE

INTEGRATED SENSOR FUSION HUB

HIGH BANDWIDTH SOC FABRIC & MEMORY SYSTEM
SIGNIFICANT DENSITY INCREASE

“Raven Ridge” die

Technology: GLOBALFOUNDRIES 14nm – 11 layer metal
Transistor count: 4.94B
Die Size: 209.78mm²

59% more transistors
16% smaller die
than prior generation “Bristol Ridge” APU

BGA Package: 25 x 35 x 1.38mm

* See footnotes for details.
INTEGRATED “VEGA” GRAPHICS

Graphics Engine
- Up to 11 Next Gen Compute Unit (NCU)
- 1 MB L2
- Flexible Geometry Engine
- 1 Draw Stream Binning Rasterizer
- 16 Pixels Units (32bpp)
- 44 Texture Units

DirectX® 12.1 Features
- Conservative Rasterization
- Raster Ordered Views
- Standard Swizzle
- Axis Aligned Rectangular Primitives

Throughput at 11 NCU
- 1200 MTris/sec @ 1200 Mhz
- Rendering 19.2 GPix/sec @1200 MHz
- 1690 FP32GFLOPS / 3379 FP16GFLOPS @ 1200 MHz
- 52.8 MTex per second @ 1200MHz
“ZEN” CPU IMPROVES VISUAL FRAME RATE

High performance “Zen” core
- Free up more power for GPU
“ZEN” WITH PRECISION BOOST 2

- Governed by CPU temperature, current, load
- Seeks highest possible frequency from environmental inputs, graceful roll-off
- Opens new boost opportunities for real-world nT workloads (e.g., games)
- 25MHz granularity
TUNE FOR THE PHASES OF VISUAL WORKLOADS
STEER POWER WHERE IT’S BEST USED

- Trade power/current based on dynamic utilization:
  - Core ↔ Core
  - CPU ↔ GPU
- On-die regulation and fine-grained frequency control enables fast, accurate frequency and voltage changes
- Fine-grained p-states (FGPS) across the IPs - continuous frequency control

3DMark® Fire Strike

Graphics phase

Physics phase

Time 35 sec 34 sec

* See footnotes for details.
"ZEN" CPU AND "VEGA" GFX CO-MANAGEMENT
WITH INFINITY FABRIC

- CPU threads feed major GPU resources: 3D engine, compute engine, and DMA engine (data fetch and writeback)
- CPU "submits" tasks, GFX "renders" or "computes"
- One coherent control and data interface to integrate and manage the full SoC
- Power budgeting based on activity and efficiency
- Enhanced flow for quiescing/powering-off CPU-GFX component
FAST DEPLOYMENT OF NEW ARCHITECTURE

MODULAR AMD INFINITY FABRIC

- Standard port definition for IP connections (SDP = Scalable Data Port)
  - Common interface definition used for CPU, GPU, I/O, multi-media hubs, display, memory controller

- Coherent HyperTransport™ transport layer
  - Builds upon generations of coherent fabric development
  - Flexible topology to adapt to diverse SoC configurations

- SDP hides complexities of coherence protocol from connected IP
“RAVEN RIDGE” INFINITY FABRIC

“Raven Ridge” Optimizations

- 32 Byte internal datapath width
- Up to 1.6GHz for bandwidth exceeding 50GB/s
- Up to 5 transfers/clock per switch
- Improved CPU latency under load, while maintaining DRAM efficiency
- Structured for multi-region power gating
- Floorplan-aware, optimized display to memory routing
QUALITY OF SERVICE FOR SMOOTH VISUAL EXPERIENCE

Three Request Classes

- Hard real time:
  - High BW (e.g., display surface refresh)
  - Low BW (e.g., audio)
- Soft real time (e.g., video playback)
- Non real time
  (e.g., typical CPU/GPU/IO requests)

Architectural Mechanisms

- Multiple virtual channels
- Priority classes (Low/Medium/High/Urgent)
- End-to-end priority escalation by VC for out of bounds conditions

Picker arbitration generally age ordered, except when younger passes older due to:

1) priority
2) VC resource availability
3) other resource such as output port busy
MEMORY BOUND PERFORMANCE OPTIMIZATION

New features and optimized SoC configuration contribute to improved memory-limited performance:

- Caching and algorithms to reduce memory requests
- Improved lossless compression usage (DCC)
- Better request ordering to reduce DRAM page conflicts and read/write turnarounds
"RAVEN RIDGE" GRAPHICS SCALING
GENERATIONAL IMPROVEMENTS FOR MEMORY BOUND GAMING PERFORMANCE

Gaming performance scaling uplift due to new AMD Vega GPU features:

- 4x larger GFX L2 cache, unified across all graphics clients
- DSBR (Draw Stream Binning Rasterizer) feature reduces bandwidth
- Improved lossless DCC memory compression

Shadows of Mordor 1920x1080 DirectX® 11

* See footnotes for details.
NEW GENERATION DISPLAY AND VIDEO CODEC ENGINE

Display Engine (DCN)
- Flexible display pipe architecture
  - Up to four 4k@60 displays
- Low power display engine with DCC, 4K2K@60hz @Vmin
- HDR support
  - From 32bpp to 64bpp surfaces
  - From sRGB to BT2020
- Higher bandwidth interfaces - HDMI 2.1, DP 1.4, HBR3
- USB-Type C with display alt-mode

Video Codec (VCN)
- Unified encode and decode engine
  - Up to 4k@60 HEVC 10b decode
  - Up to 4k@30 HEVC 8b encode
- Low power video playback – 4k@30 @Vmin
- HEVC 10b decode
- HEVC encode for superior quality skype
- VP9 decode for efficient YouTube playback
EFFICIENT POWER DELIVERY
WITH DIGITAL LOW-DROPOUT REGULATORS

- Current delivery overprovisioned for worst-case overlap between CPU and GPU
- Fine-grain LDO control allows for efficient tracking of the CPU and GFX phases, powered by a unified VDD power rail
- 1st stage: off-chip motherboard vreg
- 2nd stage: on-chip vreg with digital LDO
- Multiple digital LDO regions for CPU cores, graphics core, and sub-regions
  - Idle engine is powered off
- Allows more peak CPU/GPU current to improve boost performance

* See footnotes for details.
SYNERGISTIC POWER RAIL SHARING
WITH DIGITAL LDO REGULATORS

- Shared regulator reduces total regulator current requirements
- Less motherboard power supply footprint
- More peak CPU/GPU current to improve boost performance

CPU/GPU ICCMAX

* See footnotes for details.
ENHANCED POWER OFF STATE

CPU AND GPU

For CPU Cores
- Each core can enter CC6 power gating
- CPUOFF can lower L3 cache power when all cores in CC6

For Graphics
- Gating can power down up to 95% of the GPU
- GFXOFF can further power down GPU un-core (aka GPU monitor logic)

GFXOFF+CPUOFF=VDDOFF;
Halts System VDD Regulator
- Up to 99% residency in Windows static screen idle*

* See footnotes for details.
MORE THERMAL COMPUTE HEADROOM IN NOTEBOOKS
SKIN TEMPERATURE AWARE POWER MANAGEMENT (STAPM)

Before STAPM:
APU guard-banded to Tj~60C to meet Tskin requirements

After STAPM:
Delta between ambient and Tskin calculated based on the power/activity system components

Conceptual example of behavior
### 3DMARK® TIME SPY

<table>
<thead>
<tr>
<th>Processor Type</th>
<th>Desktop benchmark</th>
<th>Notebook benchmark</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AMD Ryzen™ 5 2400G</strong></td>
<td>1,096</td>
<td><strong>AMD Ryzen™ 7 2700U</strong></td>
</tr>
<tr>
<td>Core i5-8400</td>
<td>428</td>
<td>Core i7-8550U</td>
</tr>
<tr>
<td>Core i5-7400</td>
<td>377</td>
<td>Core i7-7500U</td>
</tr>
<tr>
<td><strong>AMD FX™ 9800P</strong></td>
<td></td>
<td>Core i7-7500U</td>
</tr>
</tbody>
</table>

*See footnotes for details.*
### Gaming on the Go

**In an ultrathin**

<table>
<thead>
<tr>
<th>Game</th>
<th>Resolution</th>
<th>API</th>
<th>Settings</th>
<th>Average FPS</th>
<th>Visual Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>League of Legends™</td>
<td>1080p</td>
<td>DirectX® 9</td>
<td>Medium</td>
<td>59</td>
<td>Good</td>
</tr>
<tr>
<td>DOTA™ 2</td>
<td>1080p</td>
<td>DirectX® 11</td>
<td>Fastest+</td>
<td>49</td>
<td>Good</td>
</tr>
<tr>
<td>Overwatch™</td>
<td>1280x720</td>
<td>DirectX® 11</td>
<td>Low 79% Render Scale</td>
<td>66</td>
<td>Good</td>
</tr>
<tr>
<td>CS:GO™</td>
<td>1080p</td>
<td>DirectX® 9</td>
<td>Medium No MSAA</td>
<td>49</td>
<td>Good</td>
</tr>
<tr>
<td>Quake® Champions</td>
<td>1280x720</td>
<td>DirectX® 11</td>
<td>High</td>
<td>43</td>
<td>Good</td>
</tr>
</tbody>
</table>

### True High-Definition

**1080p Game Performance**

<table>
<thead>
<tr>
<th>Game</th>
<th>Resolution</th>
<th>API</th>
<th>Settings</th>
<th>Average FPS</th>
<th>Visual Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Battlefield 1</td>
<td>1080p</td>
<td>Low, DX12</td>
<td>79% Render Scale</td>
<td>52</td>
<td>Good</td>
</tr>
<tr>
<td>Overwatch™</td>
<td>1080p</td>
<td>Medium</td>
<td></td>
<td>49</td>
<td>Good</td>
</tr>
<tr>
<td>Rocket League</td>
<td>1080p</td>
<td>Medium</td>
<td></td>
<td>49</td>
<td>Good</td>
</tr>
<tr>
<td>Skyrim</td>
<td>1080p</td>
<td>Medium</td>
<td></td>
<td>96</td>
<td>Good</td>
</tr>
<tr>
<td>Witcher 3</td>
<td>1080p</td>
<td>Low, Hair Works Off</td>
<td></td>
<td>31</td>
<td>Good</td>
</tr>
</tbody>
</table>

*See footnotes for details.*
Developing energy efficient processors has long been a design focus at AMD. In 2014, AMD set a bold “25x20” goal to deliver at least 25X more energy efficiency in our mobile processors by 2020. Visit AMD.com/25x20.
The true potential of the APU realized by combining “Zen” CPU with “Vega” Graphics

Data movement improvements at all levels to reduce bandwidth bottlenecks

Advances in power and thermal management provide more headroom for visual throughput