RDNA ARCHITECTURE

AMD NEXT HORIZON GAMING

7 nm
251 sqmm
10.3 Billion Transistors
GDDR 6
PCle® 4.0 Support
“NAVI”

KEY TECHNOLOGY INFLECTIONS

**PROCESS**
7NM
Faster, Smaller, Lower Power Transistors

**DRAM**
GDDR6
Cost Effective 448 GB/S of Memory Bandwidth

**INTERCONNECT BANDWIDTH**
PCle® 4.0 Support
Up to 2X Interconnect Bandwidth Of PCle®Gen3

**ARCHITECTURE**
New GFX RDNA
Designed For Gaming Performance & Efficiency
“NAVI”

Radeon™ Display Engine
New High Resolution HDR Displays
New Levels of Compression

Radeon™ Multi-Media Engine
Seamless Streaming
Improved Encoding

New Graphics RDNA Architecture
New Compute Units
Multilevel Cache
Streamlined Graphics Engine
RADEON™ DISPLAY ENGINE

FEATURING AMD RADEON FREESYNC™ TECHNOLOGY

- HDMI 2.0 & DisplayPort 1.4 HDR
- Display Stream Compression 1.2a
- Direct Read of DCC Compressed Surfaces

- Optimized for High Resolution HDR displays
  - 4K 240Hz | SINGLE CABLE | 8K 60Hz

- Optimized for Head Mounted Displays
  - Single IO connectivity

- High Fidelity Internal Color Depth
  - 30 bpp color

- Better Power Efficiency
  - Multi Plane Overlay Protocol with Low voltage mode
RADEON™ MULTIMEDIA ENGINE
SEAMLESS STREAMING

IMPROVED ENCODING
NEW HDR/WCG ENCODE (HEVC)
8K ENCODE (HEVC & VP)
40% ENCODER SPEEDUPS

<table>
<thead>
<tr>
<th>Format</th>
<th>Decode</th>
<th>Encode</th>
</tr>
</thead>
<tbody>
<tr>
<td>VP9</td>
<td>4K90</td>
<td>8K24</td>
</tr>
<tr>
<td>H.264</td>
<td>1080p600</td>
<td>1080p360</td>
</tr>
<tr>
<td></td>
<td>4K150</td>
<td>4K90</td>
</tr>
<tr>
<td>H.265</td>
<td>1080p360</td>
<td>1080p360</td>
</tr>
<tr>
<td></td>
<td>4K90</td>
<td>4K60</td>
</tr>
<tr>
<td>MPEG-4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

YouTube, twitch, NEXT GEN
RDNA
NEW GRAPHICS ARCHITECTURE

NEW COMPUTE UNIT DESIGN
Great Efficiency for Diverse Modern Gaming

MULTILEVEL CACHE HIERARCHY
Low Latency, High Bandwidth, Low Power

STREAMLINED GRAPHICS PIPELINE
Excellent Performance per Higher Clock Frequencies

DESIGNED FOR THE FUTURE OF GAMING
“NAVI” STATS
RDNA, ALL NEW ARCHITECTURE

40 RDNA Compute Units
- 80 Scalar Processors
- 2560 Stream Processors
- 160 64b Bilinear Filter units

Multilevel Cache
- 4MB L2, 512KB L1, (V$, I$, K$) L0
- 2x V$LO Load Bandwidth
- DCC Everywhere

Streamlined Graphics Engine
- Geometry Engine (4 Prim shader out, 8 Prim shader in)
- 64 Pixel Units
- 4 Asynchronous Compute Engines
- Balanced Work Distribution & Redistribution
- Designed for higher frequencies at lower power
RDNA COMPUTE UNIT

GREAT EFFICIENCY FOR DIVERSE WORKLOADS

- 2x Vector & Scalar Instruction Rate
- Single Cycle Instruction Issue
- Dual Wave Length (32\64) Modes
- Adjacent CU Resource Pooling
### RADEON™ Architectural Advances of Programmable Graphics

<table>
<thead>
<tr>
<th>Era</th>
<th>Year Range</th>
<th>Description</th>
<th>Architecture Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-2000</td>
<td></td>
<td>Fixed Function</td>
<td></td>
</tr>
<tr>
<td>1st Era</td>
<td>2001-2007</td>
<td>3D Geometry Transformation</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lighting</td>
<td></td>
</tr>
<tr>
<td>2nd Era</td>
<td>2008-2011</td>
<td>Simple VS/PS Shaders</td>
<td></td>
</tr>
<tr>
<td>3rd Era</td>
<td>2012-2018</td>
<td>Unified Shaders with VLIW</td>
<td></td>
</tr>
<tr>
<td>4th Era</td>
<td>2018-2018</td>
<td>Unified Shaders with Scalar &amp; Vector (SIMTILP Capable)</td>
<td></td>
</tr>
<tr>
<td>5th Era</td>
<td>2019-2021</td>
<td>Unified Shaders</td>
<td></td>
</tr>
</tbody>
</table>

#### Pre-2000

- Fixed Function
- 3D Geometry Transformation
- Lighting

#### 1st Era

- 3D Geometry Transformation
- Lighting

#### 2nd Era

- Simple VS/PS Shaders
- Unified Shaders with VLIW

#### 3rd Era

- Unified Shaders with VLIW
- Unified Shaders with Scalar & Vector (Fold by 4)

#### 4th Era

- Unified Shaders with Scalar & Vector (SIMTILP Capable)
- Unified Shaders

#### 5th Era

- Unified Shaders
- Unified Shaders

---

For more detailed information, please refer to the AMD Next Horizon Gaming slide deck.
RDNA COMPUTE UNIT

REDESIGNED FOR SINGLE THREADED PERFORMANCE

NEW COMPUTE UNIT

GCN COMPUTE UNIT
EXECUTION UNITS

**GCN**
- 4 SIMD16 Vector Unit
- 4 SIMD4 Special Function Unit
- 1 Shared Scalar Decode & Issue Unit
- 1 Shared Vector Decode & Issue Unit
- 256 KB VGPR

**RDNA**
- 2 SIMD32
- 2 SIMD 8 Special Function Unit
- 2 Scalar Decode and Issue Units
- 2 Vector Decode and Issue Units
- 256 KB VGPR
All work-items of a wave64 have an opportunity to do work once every 4 clocks due to hardware interleaving.

Special Function Unit alternate execution unit running at ¼ rate.

A wave from a SIMD has an opportunity to accomplish a scalar instruction once every 4 clocks.
Vector Units - All work-items of one wave32 have an opportunity to do work every clock
Special Function Unit uses 1 issue cycle and then executes in parallel
Each SIMD equipped with a scalar unit for an instruction execution every cycle
## INSTRUCTION ISSUE EXAMPLE

<table>
<thead>
<tr>
<th><strong>EXAMPLE SHADER</strong></th>
<th><strong>CYCLE</strong></th>
<th><strong>“VEGA” EXECUTION</strong></th>
<th><strong>“NAVI” WAVE32</strong></th>
<th><strong>“NAVI” WAVE64</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>s_add_i32 s0, s1, s2</td>
<td>s_add_i32 s0, s1, s2</td>
<td>s_add_i32 s0, s1, s2</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>v_mul_f32 v0, v1, s0</td>
<td>v_mul_f32 v0, v1, s0</td>
<td>v_mul_f32 v0, v1, s0</td>
</tr>
</tbody>
</table>
|                     | 5         | ... (simd busy 4 cycles) | ... (valu dependency stall on V0) | ...
|                     | 6         | ...                  | ...               | ...               |
|                     | 7         | ...                  | ...               | ...               |
|                     | 8         | v_add_f32 v5, v4, v3 | v_add_f32 v5, v4, v3 | v_add_f32 v5, v4, v3 |
|                     | 9         | ...                  | ...               | ...               |
|                     | 10        | ...                  | ...               | ...               |
|                     | 11        | ...                  | ...               | ...               |
|                     | 12        | v_sub_f32 v6, v7, v0 | v_sub_f32 v6, v7, v0 | v_sub_f32 v6, v7, v0 |
|                     | 13        | ...                  | ...               | ...               |
|                     | 14        | ...                  | ...               | ...               |
|                     | 15        | ...                  | ...               | ...               |

**SHORTEST WAVE ISSUE LATENCY**

**44% REDUCTION IN ISSUE CYCLES**

BENEFITS OF NEW WORK DISTRIBUTION & INSTRUCTION EXECUTION CHANGES

SINGLE THREADED PERFORMANCE IMPROVEMENT

WORK LOAD EXAMPLE: 64 WORK-ITEMS ALU INTENSIVE CODE

GCN

1 Wave64 ➔ SIMD16
Instruction Issue ➔ 4 clock
CU ALU ➔ 25% utilized
EFFECTIVE THROUGHPUT

RDNA

2 Wave32 ➔ 2 SIMD32
Instruction Issue ➔ 1 clock
CU ALU ➔ 100% utilized
ILP UNLOCKS UP TO 4X FASTER EXECUTION

RDNA MORE EFFECTIVELY UTILIZES THE MACHINE

Engage machine quicker by uniformly distributing work to all ALUs
Optimize efficiency and latency by preferring highest priority/oldest work
Extract program ILP and scheduling to benefit from data locality
Utilize multi-threading of waves to hide remaining latencies for throughput
RDNA SIMD UNIT

REDESIGNED FOR
SINGLE THREADED PERFORMANCE
AND
EFFECTIVE IPC

UP TO
20 Wave Controllers
WAVE 32 or WAVE64

Dedicated Instruction Issue Unit

Increased SIMD Width
SIMD 16 → SIMD 32

Improved Instruction Arbitration & Prefetch

Dedicated Scalar Unit

Full Rate 32b & Dual 16b ALU

RESULTS
REDUCED WAVE-LIFETIME & IMPROVED EFFICIENCY

Scalar Register

Stream Processors

Scalar Unit

Vector Register

Scheduler
WAVE EXECUTION
ENABLING DETERMINISTIC GAMEPLAY

MORE TOLERANT TO DIVERGENT CODES
COMPILER DRIVEN CACHE AWARE SCHEDULING

WAVE 32 EXECUTION
NATIVE ONE CLOCK EXECUTION

- Can operate in a smaller cache footprint
- Fewer resources (GPRS) required for wave launch
- Less work to hide total machine latency
- More SIMDs engaged with small workload

WAVE 64 EXECUTION
NATIVE TWO CLOCK EXECUTION

- Half wave execution improves efficiency
- Reduces exposure to ALU pipe stalls
- Enable double work item occupancy for latency hiding
Compute Unit Cooperation
Workgroup Processor

- UP TO 2X ALUs
- ACCESS TO 2X Registers
- ACCESS TO 4X Cache Bandwidth
### "NAVI" RDNA ARCHITECTURE

**Fundamental Changes in Programmable Core**

<table>
<thead>
<tr>
<th>PRE GCN</th>
<th>GCN</th>
<th>RDNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>VLIW5/VLIW4</td>
<td>Wave64 on SIMD16 (4clk issue)</td>
<td>Wave32 on SIMD32 (1clk issue)</td>
</tr>
<tr>
<td>Hard To Program for Performance</td>
<td>Easy To Program For Performance</td>
<td>Easier Achieved Performance</td>
</tr>
<tr>
<td>Complex Compiler Technology</td>
<td>Standard Compiler Techniques</td>
<td>Enables New Compiler Techniques</td>
</tr>
<tr>
<td>Per Work-Item IPC = 1.25 potential</td>
<td>Per Work-Item IPC = 0.25</td>
<td>Per Work-Item IPC = 1 potential</td>
</tr>
</tbody>
</table>
RDNA CACHE HIERARCHY
NEW MULTI LEVEL SYSTEM

- New L1 Level Cache
- Improved Bandwidth Amplification
- Reduced Latency and Power
- Reduced Congestion at L2 Level
MULTILEVEL CACHE HIERARCHY
LOW LATENCY, HIGH BANDWIDTH, LOW POWER

- Unified LLC for GFX/ACE Pipes
- Instruction Range Based Actions
- OOO between R/W, L0, L1, L2, Mem
- Reduced Latency and Power
- Reduced Data Movement
MULTILEVEL CACHE HIERARCHY
LOW LATENCY, HIGH BANDWIDTH, LOW POWER

Introduce L1 Cache Hierarchy
Double the Load Bandwidth from L0 to ALU
Reduce Latency at Each Level
Improve Effective Bandwidth

Relative Cache Latency
- L0: -21%
- L1+L2: -24%
- Memory: -7%

L0
- 16KB 32 Way
- 128KB 16 Way
- 4x64B/C

L1
- 16KB 32 Way
- 128KB 4 Way
- 64B/CLK

L2
- 16KB 32 Way
- 128KB 4 Way
- 64B/CLK

Memory
- 16x32B/C

Wave Buffers
- Wave Buffers
- SCPR
- SCPR
- SCPR
- SIMT/VP/FR
- SIMT/VP/FR

RB Cache
- 32KB 4 Way
- K Cache
- 16KB 4 Way

NEW!
MULTILEVEL CACHE HIERARCHY
DELTA COLOR COMPRESSION EVERYWHERE

Improved Color Compression Algorithm

Shaders Can Read and Write to Compressed Color Data

Display Can Now Read Compressed Color Data
STREAMLINED GRAPHICS ENGINE
IMPROVED PERFORMANCE PER CLOCK

4 Enhanced Asynchronous Compute Engines
Priority Tunneling

Centralized Geometry Processor with 4 Prim Units
Uniformly handle:
Vertex reuse, primitive assembly, reset index
Uniformly distribute pre/post tessellation work
Shader Culling - 4 Prim out, 8 Prim in

64 Pixel Units
Cache aware pixel wave packing
ENHANCED ASYNC COMPUTE
ACCELERATED PERFORMANCE
HIGH PRIORITY COMPUTE

ASYNC COMPUTE TUNNELING
PRECISE CONTROL OF OTHER WORK IN FLIGHT

Stall other pipeline launch and complete draining of other pipeline shader waves
STREAMLINED GRAPHICS PIPELINE

Improved Architectural Efficiency for Performance

Hyper-Effective Clock Gating for Power Efficiency

Reduced Levels of Logic For Higher Frequency

Leveraging “Zen” Discipline and Methodology

Performance/Clock

Normalized % Clocks Toggling

Effective Gates/Clock

See Endnotes RX-327. Data based on AMD internal testing 060119
RDNA DELIVERS

SAME POWER, GREATER PERFORMANCE

Delivered Performance

RDNA PERFORMANCE CONTRIBUTORS

- Design Frequency and Power Improvement
- 7nm Process Gains
- Performance per Clock Enhancement

See Endnotes RX-325. Data based on AMD internal testing 06/01/19.
RDNA

THE FOUNDATION FOR GREAT PRODUCTS

1.5X PERFORMANCE PER WATT

2.3X PERFORMANCE PER AREA

Absolute Performance

Power

495 mm²

14 nm "VEGA 10"

Die Size

251 mm²

7 nm "NAVI"

"Vega64" GPU
"Navi" GPU

"Vega64" GPU
"Navi" GPU
THE ALL NEW
“NAVI” GPU FAMILY FEATURES

- RDNA Architecture
- 7nm Process
- GDDR6 Memory
- PCIe® 4.0 Support
- RADEON Media Engine
- RADEON Display Engine
AMD RADEON™ GAMING IS EVERYWHERE

RDNA: EXPANDING THE RADEON UNIVERSE

- PCs
- Macs
- Consoles
- Cloud
- Mobile
ENDNOTES

RX-375
Testing done by AMD performance labs 5/23/19, using the Division 2 @ 2560x1440 settings. Performance may vary based on use of latest drivers. RX-325.

RX-377
Testing done by AMD performance labs 5/23/19, showing a geanomier of 1.25x per/clock across 30 different games @ 4K Ultra, 4K settings. Performance may vary based on use of latest drivers. RX-322.

RX-379
Testing conducted by AMD Performance Labs as of 05/30/2019 on Radeon RX 5700XT with AMD Driver 19.9.3 (13022.79466) on Intel i7-6800K, and on Radeon Vega Frontier Edition with AMD Driver 19.9.3 (13042.23183434) on Intel i7-5960X. Both systems used 2x8GB DDR4 2133MHz RAM, Asus ROG Rampage V Edition Motherboard, and Windows 10 Enterprise. Performance may vary. RX-329.

RX-358
Testing done by AMD performance labs on June 8 2019. Systems were tested with: Intel(R) Core(TM) i7-5930K CPU @ 3.50GHz (6 core) with 16GB DDR4 @ 2133 MHz using a Asus X99 E Motherboard running Windows 10 Enterprise 64-bit (Ver. 1809, build 17763.033). Using the following graphics cards: Navi (Driver 19.30_190516.1434 (C14 17840720)) with 40 compute units, versus a Vega 64 (Driver 19.4.3) with 40 compute units enabled. Running 3D Mark 11 GT1 (1280 x 720), 3D Mark 11 GT2 (1280 x 720), 3d Mark Firestrike GT1 (2560 x 1440), 3d Mark Firestrike GT2 (2560 x 1440). Unigine Heaven (1920 x 1080) , the Navi (with a die size of 251mm*2) achieved an average FPS score of 140, 136, 49, 37, and 84 respectively. Compared to the Vega 64 [with a die size of 486mm*2] which achieved 103, 113, 41, 32, and 72 respectively. RX-358.

GD-061
HEVC (H.265), H.264, and VP9 acceleration are subject to and not operable without inclusion/installation of compatible HEVC players. GD-81.

GD-127
Radeon FreeSync technology requires a monitor and AMD Radeon™ graphics, both with FreeSync support. See www.amd.com/freesync for complete details. Confirm capability with your system manufacturer before purchase. GD-127.

DISCLAIMER
The information contained herein is for informational purposes only, and is subject to change without notice. While every precaution has been taken in the preparation of this document, it may contain technical inaccuracies, omissions or typographical errors, and AMD is under no obligation to update or otherwise correct this information. Advanced Micro Devices, Inc. makes no representations or warranties with respect to the accuracy or completeness of the contents of this document, and assumes no liability of any kind, including the implied warranties of non-infringement, merchantability or fitness for particular purposes, with respect to the operation or use of AMD hardware, software or other products described herein. No license, including implied or arising by estoppel, to any intellectual property rights is granted by this document. Terms and limitations applicable to the purchase or use of AMD's products are set forth in a signed agreement between the parties or in AMD's Standard Terms and Conditions of Sale.

ATTRIBUTION
© 2019 Advanced Micro Devices, Inc. All rights reserved. AMD, the AMD Arrow logo, Radeon, Ryzen, FreeSync, and combinations thereof are trademarks of Advanced Micro Devices, Inc. in the United States and/or other jurisdictions. Other names are for informational purposes only and may be trademarks of their respective owners.