“ZEN 2”

David Suggs | Dan Bouvier

Michael Clark, Kevin Lepak, Mahesh Subramony

August 19, 2019
THE PATH TO “ZEN 2”

CHALLENGES

- Meet multiple markets – server to mobile
- Ultimate performance
- Energy efficient performance

TECHNOLOGY

- Enable scale and power
- Balancing complexity of new technology node with time-to-market

PLATFORM

- Chiplets to deliver right technology to the needs
- Upgrade IO to meet system demands
- Compatibility
“ZEN 2”

MICROARCHITECTURAL HIGHLIGHTS

- New TAGE branch predictor
- 2x op cache capacity
- Reoptimized L1I cache
- 3rd address generation unit
- 2x FP data path width
- 3x L1 load+store bandwidth
- 2x L3 capacity
- Improved prefetch throttling
- 2 threads per core (SMT) carried forward

15% IPC IMPROVEMENT FROM “ZEN” TO “ZEN 2”
MAJOR EFFICIENCY IMPROVEMENTS

- Improved branch prediction accuracy
- Higher op cache hit rate
- New integer scheduler algorithms
- Clock and data gating improvements
- Low-power design methodology
## Designed for Security

### Hardware Optimized Security Mitigations

<table>
<thead>
<tr>
<th>Spectre v2*</th>
<th>Indirect branch target injection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectre v4**</td>
<td>Speculative store bypass</td>
</tr>
</tbody>
</table>

### Secure Microarchitecture

<table>
<thead>
<tr>
<th></th>
<th>Hardware permission checks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Permissions are checked prior to consuming data</td>
<td></td>
</tr>
</tbody>
</table>

### Secure Virtualization

<table>
<thead>
<tr>
<th></th>
<th>SEV-ES</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEV with encrypted state</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>More SEV Keys</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEV with up to 509 encrypted guests (was 15)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>SEV-VTE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEV with virtual transparent encryption</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>GMET</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guest mode execute trap. Hypervisor can trap guest supervisor-mode execution of pages for integrity checking</td>
<td></td>
</tr>
</tbody>
</table>
# Extending the Architecture

<table>
<thead>
<tr>
<th>Feature</th>
<th>Notes</th>
<th>“Zen”</th>
<th>“Zen 2”</th>
</tr>
</thead>
<tbody>
<tr>
<td>x2APIC</td>
<td>APIC extension for high core count system support</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>QOS</td>
<td>Quality-of-service monitoring/enforcement of L3 cache occupancy and memory bandwidth</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>UMIP</td>
<td>User mode instruction prevention (CR4.UMIP controls access to SGDT, SIDT, SDLT, STR, and SMSW)</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>CLWB</td>
<td>Non-volatile memory enhancement (cache-line writeback)</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>WBNOINVD</td>
<td>Cleans caches, but does not invalidate</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>RDPRU</td>
<td>Read processor register at a user level (currently MPERF and APERF)</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

- ❄️ AMD Exclusive
PREDICTION, FETCH, AND DECODE
GREATER ACCURACY AND HIGHER CAPACITY

<table>
<thead>
<tr>
<th>IMPROVED BRANCH PREDICTION</th>
<th>“ZEN”</th>
<th>“ZEN 2”</th>
</tr>
</thead>
<tbody>
<tr>
<td>L2 Predictor</td>
<td>Perceptron</td>
<td>TAGE</td>
</tr>
<tr>
<td>2x L0 BTB</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>2x L1 BTB</td>
<td>256</td>
<td>512</td>
</tr>
<tr>
<td>1.75x L2 BTB</td>
<td>4K</td>
<td>7K</td>
</tr>
<tr>
<td>2X Indirect Target Array</td>
<td>0.5K</td>
<td>1K</td>
</tr>
<tr>
<td>30% Lower Mispredict Rate Target</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OPTIMIZED L1I CACHE</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2x Associativity</td>
<td>64K</td>
<td>32K</td>
</tr>
<tr>
<td>Improved Utilization</td>
<td>4-way</td>
<td>8-way</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2X OP CACHE (FUSED INSTR)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2K</td>
<td>4K</td>
</tr>
<tr>
<td>Better Instruction Fusion</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Better Effective Throughput</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

REDIRECT FROM DISPATCH/EXECUTE

- Next PC
- L0/L1/L2 TLB
- L0/L1/L2 BTB
- Return Stack
- Indirect Target Array
- 32K L1I Cache
- Micro-Tags
- Decode
- Op Cache
- Micro-Op Queue
- Microcode Rom
- Dispatch
- 6 MICRO-OP DISPATCH
## INTEGER EXECUTE

### WIDER ISSUE AND DEEPER WINDOW

<table>
<thead>
<tr>
<th></th>
<th>“ZEN”</th>
<th>“ZEN 2”</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WIDER ISSUE</strong></td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>ALUs</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>AGUs</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

### DEEPER WINDOW

<table>
<thead>
<tr>
<th>Feature</th>
<th>“ZEN”</th>
<th>“ZEN 2”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bigger ALU Scheduler</td>
<td>4x14</td>
<td>4x16</td>
</tr>
<tr>
<td>More Unified AGU Scheduler</td>
<td>2x14</td>
<td>1x28</td>
</tr>
<tr>
<td>Bigger Register File</td>
<td>168</td>
<td>180</td>
</tr>
<tr>
<td>Bigger Reorder Buffer</td>
<td>192</td>
<td>224</td>
</tr>
</tbody>
</table>

### IMPROVED SMT FAIRNESS FOR ALU AND AGU SCHEDULERS

- ✔️
## FLOATING-POINT/VECTOR EXECUTE

**DOUBLE WIDE VECTORS**

<table>
<thead>
<tr>
<th>Feature</th>
<th>“ZEN”</th>
<th>“ZEN 2”</th>
</tr>
</thead>
<tbody>
<tr>
<td>AVX256 INSTRUCTION SUPPORT</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>2X WIDTH DATA PATH</td>
<td>128b</td>
<td>256b</td>
</tr>
<tr>
<td>EDC Management</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>2X WIDTH VECTOR REGISTER FILE</td>
<td>128b</td>
<td>256b</td>
</tr>
<tr>
<td>2X WIDTH LOADS (2)</td>
<td>128b</td>
<td>256b</td>
</tr>
<tr>
<td>2X WIDTH STORES (1)</td>
<td>128b</td>
<td>256b</td>
</tr>
<tr>
<td>IMPROVED DOUBLE-PRECISION MULTIP. LATENCY</td>
<td>4 cyc</td>
<td>3 cyc</td>
</tr>
</tbody>
</table>

**Diagram:**
- 4 MICRO-OP DISPATCH
- 64-Entry NSQ
- 32-Entry Scheduler
- 160-Entry Vector Register File
- Forwarding MUXes
- FMA
- FADD
- FMA
- FADD
- 224-Entry Reorder Buffer
- LDCVT
- 256B LOADS
- INT TO FP
- FP TO INT, STORE
LOAD/STORE AND L1D CACHE
MORE THROUGHPUT AND BIGGER STRUCTURES

<table>
<thead>
<tr>
<th>Feature</th>
<th>“ZEN”</th>
<th>“ZEN 2”</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIGGER STORE QUEUE</td>
<td>44</td>
<td>48</td>
</tr>
<tr>
<td>BIGGER, BETTER L2 DT LB</td>
<td>1.5K</td>
<td>2K</td>
</tr>
<tr>
<td>1G Page Support (as 2M)</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Lower Latency</td>
<td>8 cyc</td>
<td>7 cyc</td>
</tr>
</tbody>
</table>

32KB, 8-WAY L1D CACHE

- 2X Width Reads (2)          | 128b  | 256b    |
- 2X Width Writes (1)         | 128b  | 256b    |
- 3X Load + Store Bandwidth   | 32B/clk | 96B/clk |

IMPROVED WRITE-COMBINING BUFFER PERFORMANCE ✔
IMPROVED PREFETCH THROTTLING ✔
“ZEN 2” SOLUTIONS

3rd Generation Ryzen™ ("Matisse")
2nd Generation Epyc™ ("Rome")
**CACHE HIERARCHY AND CPU COMPLEX**

- **2X L1D CACHE LOAD/STORE BANDWIDTH vs. “ZEN”**
  - 32B Per Cycle Everywhere
- **2X L3 CACHE UP TO 16MB PER CCX, 2CCXs per CCD**
AMD GOAL
DEPLOY LEADERSHIP PERFORMANCE SIMULTANEOUSLY INTO MULTIPLE MARKETS

CHALLENGES
- Higher cost per mm²
- Complex analog/IO elements do not scale well and can be challenging to port
- Wide range of diverse markets

SOLUTION | CHIPLET STRATEGY
- Small die improve yield
- Analog / IO IP in 12nm technology to reduce cost and complexity
- Enables product configurability

Next Generation Highly Optimized “Zen 2” Core
Leading Edge 7nm Technology for Density and Power Efficiency
Timely Delivery Of Products
REVOLUTIONARY CHIPLET DESIGN

Each IP in its Optimal Technology

Infinity Fabric™ Enables Modularity (MCM),
Scaling (CCD Count)

Optimized I/O Die Enables Common Latency to All Cores/Caches

3.8 Billion FETs, 74 mm²

12nm

Server IO Die
8.34 Billion FETs, 416 mm²

“Rome”

7nm

12nm

Client IO Die
2.09 Billion FETs, 125 mm²

“Matisse”

X570 Chipset
PUTTING IT TOGETHER
PERFORMANCE DESKTOP “MATISSE”

- A Desktop SoC and a Chipset
  - IOD combined with CCD(s) form the CPU
  - Standalone IOD re-purposed as Chipset
- Leading I/O
  - 48GB/s native PCIe™ BW
  - 4 USB 3.1 10Gb/s ports
- Memory BW
  - 51.2 GB/s Memory BW
  - Dual Channel DDR4 3200 MT/s
- Overclocking
  - Improved Memory overclocking (Phy, Package)
  - De-coupled various IO-die clocks for flexibility
- AM4 Platform Longevity
  - Compatible with AM4 platform
SIGNIFICANT DESKTOP PERFORMANCE IMPROVEMENT

Application Performance
In Power Efficient TDP

- Compute heavy workloads benefit from power efficient design
- Content Creation, Rendering benchmarks see large gains
- Additional cores in 3900X deliver substantial compute power

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Ryzen 7 2700X</th>
<th>Ryzen 7 3800X</th>
<th>Ryzen 9 3900X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cinebench R20</td>
<td>+24%</td>
<td>+67%</td>
<td>+80%</td>
</tr>
<tr>
<td>POV-Ray 3.7</td>
<td>-16%</td>
<td>-35%</td>
<td>-34%</td>
</tr>
<tr>
<td>Adobe Premiere Pro CC</td>
<td>-18%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Handbrake 1.1.1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
SIGNIFICANT DESKTOP PERFORMANCE IMPROVEMENT

Gaming Performance

- Significant generational performance improvement
- Increased L3 (up to 64MB) reduces effective memory latency

Ryzen 7 2700X
Ryzen 7 3800X

<table>
<thead>
<tr>
<th>Game</th>
<th>Ryzen 7 2700X</th>
<th>Ryzen 7 3800X</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS:GO</td>
<td>+34%</td>
<td>+30%</td>
</tr>
<tr>
<td>League of Legends</td>
<td>+22%</td>
<td>+15%</td>
</tr>
<tr>
<td>PUBG</td>
<td>+21%</td>
<td>+11%</td>
</tr>
<tr>
<td>Overwatch</td>
<td>+21%</td>
<td>+15%</td>
</tr>
<tr>
<td>DOTA 2</td>
<td>+15%</td>
<td>+11%</td>
</tr>
<tr>
<td>GTA V</td>
<td>+11%</td>
<td>+11%</td>
</tr>
</tbody>
</table>
APPLICATION BENEFITS FROM PCIE™ GEN4

IO Performance
- Up to 2x PCIe BW vs prior Ryzen generation

Storage Performance
- Large Block Sequential accesses are severely limited by link speeds

3DMark® PCIe Feature Test
- Vertex animation (game VFX) is sensitive to bus bandwidth, allowing significant upside

DaVinci Resolve
- Bus bandwidth is a significant limiting factor for non-linear editing (NLE) performance

3DMARK PCIe FEATURE TEST
- Bus Bandwidth (GBps)
- Frames Per Second

CRYSTALDISKMARK SEQUENTIAL PERFORMANCE
- NVMe Write (GB/s)
- NVMe Read (GB/s)

- AMD Ryzen™ 7 3800X Radeon™ RX 5700 XT (PCIe® Gen3)
- AMD Ryzen™ 7 3800X Radeon™ RX 5700 XT (PCIe® Gen4)
PUTTING IT TOGETHER
SERVER “ROME”

- ~1000mm² cumulative area in leading-edge 7nm/12nm design
- Up-to-64 cores (8 CCDs) and 256MB L3-cache
- Improved NUMA
  - All memory and I/O hosted by single die
  - Single IOD results in effective latency improvements
- Class-Leading I/O and Memory Bandwidth
  - 8 Memory Channels 64b DDR4-3200 204.8 GB/s Bandwidth
  - PCIe™ Gen 4
- Platform Longevity; Compatible with Prior Generation EPYC™
- Advanced Platform Security Features: Expanded SEV keys, SEV-IO, SEV-ES
TWO SOCKET LEADERSHIP
2S INTEL® XEON® VS. 2S AMD EPYC™ SPEC CPU® 2017 PERFORMANCE

- Up to 128 cores (8 CCDs) and 256MB L3-cache per socket
“ROME” LEADERSHIP PERFORMANCE

WITH REAL WORLD RESULTS

ENGINEERING SIMULATIONS
- get it right
- Altair
- UP TO 58%
  HIGHER PERFORMANCE

STRUCTURAL ANALYSIS
- UP TO 72%
  HIGHER PERFORMANCE

FINITE ELEMENT ANALYSIS
- UP TO 79%
  HIGHER PERFORMANCE

MANUFACTURING
- SIEMENS
- UP TO 95%
  HIGHER PERFORMANCE

FLUID DYNAMICS
- ANSYS
- UP TO 95%
  HIGHER PERFORMANCE

SEE ENDNOTES ROM-63, ROM-56, ROM-49, ROM-42, ROM-70
CONCLUSION

“ZEN 2” Delivers Performance Uplift
- 15% IPC
- Up to 2x instructions per unit energy

Chiplet Deployment And Partitioning
- Enable efficient targeting of technology – performance/power/cost
- Faster deployment of product stack - Client CPU, Server CPU, Chipset

Resulting In Industry Leading Products
- 3rd generation Ryzen (“Matisse”)
- 2nd generation EPYC (“Rome”)
ENDNOTES

Slide 3: Claim: 15% IPC uplift
- AMD "Zen 2" CPU-based system scored an estimated 15% higher than previous generation AMD "Zen" based system using estimated SPECInt® base2006 results. SPEC and SPECint are registered trademarks of the Standard Performance Evaluation Corporation. See www.spec.org. GD:141

Slide 4: Claim: Up to 2x Instructions per unit energy
- Testing conducted by AMD Performance Labs as of 7/12/2019 with 2nd Generation Ryzen and 3rd Generation Ryzen engineering samples using estimated SPECInt® base2006 results. PC manufacturers may vary configurations yielding different results. SPEC and SPECint are registered trademarks of the Standard Performance Evaluation Corporation. See www.spec.org.

Slide 5
- * AMD recommended settings for these security features can be found at https://developer.amd.com/wp-content/resources/Architecture_Guidelines_Update_Indirect_Branch_Control.pdf.
- ** AMD recommended settings for these security features can be found at https://developer.amd.com/wp-content/resources/124441_AMD64_SpeculativeStoreBypassDisable_Whitepaper_final.pdf.

Slide 15
- AMD's product warranty does not cover damages caused by overclocking, even when overclocking is enabled via AMD hardware and/or software. GD:26

Slide 16, 17
See table at right for test system configurations.

- Gaming: Testing by AMD performance labs using an AMD Ryzen™ 9 3900X, AMD Ryzen™ 7 3800X and AMD Ryzen™ 7 2700X. All games tested at 1920x1080 with maximum in-game quality preset. Results may vary.
- Unless otherwise noted in the legend(s) of the chart, all performance analyses conducted and published throughout Computex, Next Horizon, Gaming, E3, and the Ryzen™ Processor Reviewer’s Guide were performed on the following system configurations in an air-conditioned climate of 70°F/21°C.

Slide 18
- 2x PCIe BW over prior Ryzen generation is based on theoretical maximum bandwidth of PCIe Gen3 vs PCIe Gen4 speeds.
ENDNOTES (CONT.)

Slide 19, 20

- Faster D2D interconnect and assembly of all memory controllers and PHYs on a single IOD, improves effective memory latency in NPS1 over prior generation EPYC™.

- Each AMD EPYC processor has 8 memory channels. Each Intel Xeon Scalable processor has 6 memory channels. \(8 - 6 = 2 \div 6 = 0.33\) AMD EPYC has 33% more memory bandwidth. Class based on industry-standard pin-based (LGA) X86 processors. EPYC-06.

Slide 20:

- Slide represents both published and estimated SPECrate®2017_int_peak performance. Estimates as of July 3, 2019 for AMD EPYC 48C, 32C and 8C processors using computer modeling of preproduction parts and SPECrate®2017_int_peak internal testing results. Results may vary with production silicon testing. Published results for EPYC 64C processor as of August 7, 2019: https://spec.org/cpu2017/results/res2019q3/cpu2017-20190722-16242.html. Intel results as of June 2019: Xeon Platinum:

Slide 21:

- Based on AMD internal testing of ANSYS FLUENT 19.1, Im6000_16m benchmark, as of July 17, 2019 of a 2P EPYC 7742 powered reference server versus a 2P Intel Xeon Platinum 8280 powered server. Results may vary. ROM-42

- Based on AMD internal testing of LS-DYNA R9 3.0, neon benchmark, as of July 17, 2019 of a 2P EPYC 7742 powered reference server versus a 2P Xeon Platinum 8280 powered server. Results may vary. ROM-49

- Based on AMD internal testing of Altair RADIOSS 2018, T10M benchmark, as of July 17, 2019 using a 2P EPYC 7742 powered reference server versus a 2P Xeon Platinum 8280 powered server. Results may vary. ROM-56

- Based on AMD internal testing of ESI VPS 2018.0, NEON4m benchmark, as of July 17, 2019 using a 2P EPYC 7742 powered reference server versus a 2P Xeon Platinum 8280 powered server. Results may vary. ROM-63

- Based on AMD internal testing of Siemens PLM Star-CCM+ 14.02.009, kcs_with_physics benchmark, as of July 17, 2019 using a 2P EPYC 7742 powered reference server versus a 2P Xeon Platinum 8280 powered server. Results may vary. ROM-70