Designing a High-Performance Core

MICHAEL CLARK
CAUTIONARY STATEMENT

This presentation contains forward-looking statements concerning Advanced Micro Devices, Inc. (AMD) including, but not limited to, the features, functionality, availability, timing, expectations and expected benefits of AMD future products, including Ryzen™ 5000 Series CPUs and Socket AM4, which are made pursuant to the Safe Harbor provisions of the Private Securities Litigation Reform Act of 1995. Forward-looking statements are commonly identified by words such as "would," "may," "expects," "believes," "plans," "intends," "projects" and other terms with similar meaning. Investors are cautioned that the forward-looking statements in this presentation are based on current beliefs, assumptions and expectations, speak only as of the date of this presentation and involve risks and uncertainties that could cause actual results to differ materially from current expectations. Such statements are subject to certain known and unknown risks and uncertainties, many of which are difficult to predict and generally beyond AMD's control, that could cause actual results and other future events to differ materially from those expressed in, or implied or projected by, the forward-looking information and statements. Investors are urged to review in detail the risks and uncertainties in AMD's Securities and Exchange Commission filings, including but not limited to AMD's Quarterly Report on Form 10-Q from the quarter ending on June 26, 2021.
OUR “ZEN” JOURNEY

2017

“ZEN” / “ZEN+”
- Up to 4.35GHz max boost\(^4\)
- +52% IPC\(^1\)
- 4-core complex
- Up to 8MB L3 per complex
- SMT enabled
- New boost algorithms
- 14nm/12nm

2021

“ZEN 2”
- Up to 4.7GHz max boost
- +15% IPC\(^2\)
- 4-core complex
- Up to 16MB L3 per complex
- Chiplet design
- FP-256
- 7nm

“ZEN 3”
- Up to 4.9GHz max boost
- +19% IPC\(^3\)
- New 8-core complex
- Up to 32MB L3 per complex
- AMD 3D V-Cache support
- Doubled INT8 throughput
- 7nm
### 5 Tenets of Core Performance

<table>
<thead>
<tr>
<th>IPC</th>
<th>POWER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Architectural Efficiency</td>
<td>Environment Integration</td>
</tr>
<tr>
<td></td>
<td>Energy per operation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AREA</th>
<th>FREQUENCY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core Density Cost</td>
<td>Max Effective</td>
</tr>
</tbody>
</table>

**SOFTWARE**

- ISA
- Security
2 Socket Server Performance

Total Optimized SPECint®_rate_base2006 (est.) Performance

~37% CAGR (est.)

CAGR based on AMD internal Analysis, October 2021

See end note MLN-158
Adding cores is balanced with memory bandwidth to “feed the core”

Historically sits in a range of 3 to 6 GB/s

*http://www.spec.org
ZEN 3 SECURITY & ISA FEATURES
AMD INFINITY GUARD
NEW LAYERS OF SECURITY FOR TENANTS IN THE CLOUD

SEV
SECURE ENCRYPTED VIRTUALIZATION
Encrypt Each VM with Unique Keys

SEV-ES
ENCRYPTED STATE
VM Integrity with Protected CPU Registers

SEV-SNP
SECURE NESTED PAGING
Hardware Protection Against Malicious Hypervisors
# ISA ENHANCEMENTS

<table>
<thead>
<tr>
<th>FEATURE</th>
<th>NOTES</th>
<th>CLIENT SoCs</th>
<th>SERVER SoCs</th>
</tr>
</thead>
<tbody>
<tr>
<td>256-bit VAES/VPCLMULQDQ</td>
<td>256-bit instruction extensions for accelerating encryption / decryption algorithms</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Memory Protection Keys for Users</td>
<td>Application control for access-disable and write disable settings w/o TLB management</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>CET Shadow Stack</td>
<td>Helps protect against ROP (Return Oriented Programming) attacks by mirroring return addresses on a shadow stack, requires OS and/or hypervisor enablement</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SEV-ES Enhancements</td>
<td>Interrupt injection restrictions: Limit types of interrupts/exceptions that a (malicious) hypervisor may inject into an SEV-ES guest Debug registers added to swapped state</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Secure Nested Paging</td>
<td>Builds on confidentiality established by encryption of VM memory and VM registers in SEV/SEV-ES to add integrity protection features to help protect against malicious hypervisors including protection against replay/corruption/remapping attacks</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>INVLPGB</td>
<td>New instruction, use instead of inter-core interrupts to broadcast page invalidates, requires OS and/or hypervisor enablement</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Process Context ID (PCID)</td>
<td>Process tags in TLB to reduce flush requirements</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
“ZEN 3” OBJECTIVES

PERFORMANCE
- Deliver another landmark increase in 1T performance through IPC and frequency
- Unify cores and cache in a contiguous 8-core complex to improve effective latency
- Provide scale-out performance for servers, data-centers and super-computers

NEW CAPABILITIES
- Introduce new ISA extensions
- Expanded security features
- Support for AMD 3D V-Cache integration

PLATFORM
- Support for scaling and energy efficiency
- Socket compatibility
“ZEN 3” OVERVIEW

2 THREADS PER CORE (SMT)
STATE-OF-THE-ART BRANCH PREDICTOR

CACHES
- l-cache 32k, 8-way
- Op-cache, 4K instructions
- D-cache 32k, 8-way
- L2 cache 512k, 8-way

DECODE
- 4 instructions / cycle decode or 8 ops from Op-cache
- 6 ops / cycle dispatched to Integer or FP

EXECUTION CAPABILITIES
- 4 integer units
- Dedicated branch and store data units
- 3 address generations per cycle
- 2 256-bit FP multiply accumulate / cycle

3 MEMORY OPS PER CYCLE

TLBs
- L1 64 entries I & D, all page sizes
- L2 512 I, 2K D, everything but 1G
FETCH / DECODE

REDUCED LATENCIES
- Lower mispredict penalty
- No “bubble” on most taken branch predictions

IMPROVED BRANCH PREDICTION
- TAGE branch predictor
- Redistributed BTBs for better prediction latency
  - L1 BTB, 1024 entries
  - L2 BTB, 6.5K entries
- Larger 1.5K indirect target array (ITA)

OPTIMIZED 32KB, 8-WAY L1I CACHE
- Improved prefetching
- Improved utilization

STREAMLINED OP-CACHE
- Faster sequencing of Op-cache fetches
- Finer-grained switching of Op-cache / I-cache pipes

FASTER FETCH, ESPECIALLY FOR BRANCHY AND LARGE FOOTPRINT CODE
INT EXECUTION

- New distributed scheduler organization
- Lower latencies for some instructions
- Larger out-of-order window
- 10 issue per cycle, up from 7

<table>
<thead>
<tr>
<th>RESOURCE</th>
<th>“ZEN 2”</th>
<th>“ZEN 3”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integer issue width</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>Integer register file</td>
<td>180</td>
<td>192</td>
</tr>
<tr>
<td>Integer scheduler</td>
<td>92</td>
<td>96</td>
</tr>
<tr>
<td>ROB</td>
<td>224</td>
<td>256</td>
</tr>
</tbody>
</table>

LOWER LATENCIES AND LARGER STRUCTURES TO EXTRACT ILP FOR FEEDING THE EXECUTION ENGINES
WIDER INTEGER EXECUTION

PICK BANDWIDTH IS INCREASED

- Still four “ALU” and three “AGU” execution units
  - But adds branch and store data capabilities
  - Up to 10 integer ops picked per cycle
- No increase in register file write ports or bypass network inputs
- Shared ALU/AGU schedulers allow for balanced use across workloads

DELIVERING WIDER EXECUTION RESOURCES IN A POWER- AND AREA-EFFICIENT MANNER
FP EXECUTION

- Increased Dispatch Bandwidth (6-wide)
- Larger Scheduler
- Separate F2I/Store Units
- Faster 4-cycle FMAC
- Doubled INT8 throughput

<table>
<thead>
<tr>
<th>RESOURCE</th>
<th>“ZEN 2”</th>
<th>“ZEN 3”</th>
</tr>
</thead>
<tbody>
<tr>
<td>FP issue width</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>FADD / FMUL / FMA latency</td>
<td>3/3/5</td>
<td>3/3/4</td>
</tr>
<tr>
<td>FP scheduler</td>
<td>36</td>
<td>64</td>
</tr>
</tbody>
</table>

LOWER LATENCIES AND LARGER STRUCTURES TO EXTRACT ILP FOR FEEDING THE EXECUTION ENGINES
LOAD/STORE

LARGER STORE QUEUE (64, UP FROM 48)

PREFETCH IMPROVEMENTS
- More consistent prefetch on page crossing
- Better L1/L2 cache prefetch coordination
- MSR control of prefetch enablement (server)

MORE LOADS / STORES PER CYCLE
- 3 loads per cycle (max 2 if 256b)
- 2 stores per cycle (max 1 if 256b)
- Max 3 total memory ops

2K ENTRY L2 DTLB
- 6 page table walkers for misses

32KB, 8-WAY L1 DATA CACHE

FASTER COPY OF SHORT STRINGS

BETTER PREDICTION OF STORE-TO-LOAD DEPENDENCIES

LARGER STRUCTURES AND BETTER PREFETCHING TO EXTRACT ILP FOR FEEDING WIDER EXECUTION
### MAJOR CHANGES VS. “ZEN 2”

#### “ZEN 2”

- **Int**: 10K L1 Cache (2 x 5K)
- **Decode**: 4 Instructions
- **Micro-Op Queue**: 3 Dispatch Ops
- **Op Cache**: 8 Trued Instructions

#### FRONT-END ENHANCEMENTS

- **2X Larger L1 BTB (1024)**
- Improved branch predictor bandwidth
- “No-bubble” branch prediction
- Faster recovery from mispredict
- Faster sequencing of Op-cache fetches
- Quicker switching of Op-cache pipes

#### “ZEN 3”

- **Int**: 6 Instructions/Cycle
- **Op Queue**: 8 Micro Ops/Cycle

#### LOAD / STORE

- Higher load bandwidth (+1)
- Higher store bandwidth (+1)
- More flexibility in load/store ops
- Improved memory dependence detection
- TLB: 6 table walkers (+4)

#### EXECUTION

- **Int**: Dedicated Branch / St-data pickers
- **FP/Int**: Reduced latency for select ops
- **FP**: 6-wide dispatch and issue (+2)
- **FP**: Faster FMAC (-1 cycle)
DOUBLE DIGIT IPC GAIN - AGAIN

“ZEN 3” 19% IPC UPLIFT FOR PCs*

GEOMEAN OF 25 WORKLOADS
(Fixed 4GHz Frequency, 8 Cores)

+19%

“ZEN 3” PERFORMANCE CONTRIBUTORS

- Cache Prefetching
- Execution Engine
- Branch Predictor
- Micro-op Cache
- Front End
- Load/Store

* See endnote RS003.
"ZEN 3" IPC UPLIFT

GEOMEAN: +19% VS. "ZEN 2"

"ZEN 3" IMPROVEMENTS VS. "ZEN 2"

FIXED 4GHZ, 8 CORES

* See endnote RSK-003.
BEYOND THE CORE

AMD RYZEN™ 5000 SERIES SOC
ARCHITECTURE CHANGES
“ZEN 2” LAYOUT

2X L3 Cache Directly Accessible Per Core

“ZEN 3” LAYOUT

Accelerates Core and Cache Communication for Gaming

32MB L3 CACHE

Reduction in Effective Memory Latency
New Bi-Directional Ring Bus

High Bandwidth Low Latency

32 Bytes Each Direction
“ZEN 3” CACHE HIERARCHY
(8-CORE CCD)

- Fast private 512K L2 cache
- High bandwidth interfaces at all levels
- L3 is filled from L2 victims (i.e., mostly exclusive)
- L2 tags duplicated in L3 for probe filtering and fast cache transfer
- 64 outstanding misses supported from L2 to L3 per core
- 192 outstanding misses supported from L3 to memory
- L3 shared among all 8 cores in the complex
- Support for AMD 3D V-Cache
AMD 3D V-CACHE
192M L3 PROTOTYPE

- Zen 3 base CCD design includes 32M L3 cache
- Increased to 96M per CCD with 64M AMD 3D V-Cache
- Enabled by Through Silicon Vias on CCD
- Direct copper-to-copper bond

15% FASTER GAMING ON AVERAGE*

* See endnote RSX-078.
AMD DRIVING HIGH PERFORMANCE COMPUTING INTO THE FUTURE
ENDNOTES

CD-122: “Zen” is a codename for AMD architecture and is not a product name.

CD-150: Max boost for AMD Ryzen processors is the maximum frequency achievable by a single core on the processor running a bursty single-threaded workload. Max boost will vary based on several factors, including, but not limited to: thermal paste; system cooling; motherboard design and BIOS, the latest AMD chipset driver; and the latest OS updates.

R5K-003: Testing by AMD performance labs as of 09/01/2020. IPC evaluated with a selection of 25 workloads running at a locked 4GHz frequency on 8-core “Zen 2” Ryzen 7 3800XT and “Zen 3” Ryzen 7 5800X desktop processors configured with Windows® 10, NVIDIA GeForce RTX 2080 Ti (451.77), Samsung 860 Pro SSD, and 2x8GB DDR4-3600. Results may vary.

R5K-078 Testing by AMD performance labs as of April 22, 2021 based on the average FPS of 32 PC games at 1920x1080 with the High image quality preset using an AMD Ryzen™ 9 5900X processor vs 12-Core 3D Chiplet Prototype. Results may vary.

RZ3-24: Based on AMD Labs testing in May 2019, an AMD “Zen 2”-based system configured with a “Matisse” B0 sample, AMD Reference Mobo, AMD Reference Cooler, 4x8GB DDR4-2667 RAM, Ubuntu 19.04, and GeForce GTX 1080 GPU vs. a similarly configured “Summit Ridge” B2 sample scored an estimated 15% higher using estimated SPECint®_base2006 results. SPEC and SPECint are registered trademarks of the Standard Performance Evaluation Corporation. See www.spec.org

CD-108: Generational IPC uplift for the “Zen” architecture vs. “Piledriver” architecture is +52% with an estimated SPECint_base2006 score compiled with GCC 4.6 -O2 at a fixed 3.4GHz. Generational IPC uplift for the “Zen” architecture vs. “Excavator” architecture is +64% as measured with Cinebench R15 1T, and also +64% with an estimated SPECint_base2006 score compiled with GCC 4.6 -O2, at a fixed 3.4GHz. System config: AMD reference motherboard(s), AMD Radeon™ R9 290X CPU, 8GB DDR4-2667 (“Zen”)/8GB DDR3+2133 (“Excavator”)/8GB DDR3+1866 (“Piledriver”), Ubuntu Linux 16.x (SPECint_base2006 estimate) and Windows® 10 x64 R51 (Cinebench R15). SPECint_base2006 estimates: “Zen” vs. “Piledriver” (31.5 vs. 20.7 +52%), “Zen” vs. “Excavator” (31.5 vs. 19.2 +64%). Cinebench R15 1T scores: “Zen” vs. “Piledriver” (135 vs. 79 both at 3.4G +76%), “Zen” vs. “Excavator” (160 vs. 97.5 at 4.0G +64%).

MLN-158: SPECint®_rate_base2006 estimated scores for both AMD and Intel servers are based on two estimation methods. For servers with published SPECint®_rate_base2006 results, estimates match the published scores. For servers which do not have published SPECint®_rate_base2006 results (newer servers), estimates of SPECint®_rate_base2006 are calculated by taking the server’s published SPECrate®2017_int_base and scaling it by the ratio between SPECint®_rate_base2006 and SPECrate®2017_int_base on a different server where both scores have been published. SPEC CPU® 2006 is a retired benchmark and SPEC® is no longer reviewing or publishing SPEC CPU 2006 results. SPEC and SPEC CPU are registered trademarks of the Standard Performance Evaluation Corporation. See more information about SPEC CPU at www.spec.org
COPYRIGHT AND DISCLAIMER

©2021 Advanced Micro Devices, Inc. All rights reserved.

AMD, the AMD Arrow logo, EPYC, Ryzen, Infinity fabric, and combinations thereof are trademarks of Advanced Micro Devices, Inc. Other product names used in this publication are for identification purposes only and may be trademarks of their respective companies.

The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions, and typographical errors. The information contained herein is subject to change and may be rendered inaccurate for many reasons, including but not limited to product and roadmap changes, component and motherboard version changes, new model and/or product releases, product differences between differing manufacturers, software changes, BIOS flashes, firmware upgrades, or the like. Any computer system has risks of security vulnerabilities that cannot be completely prevented or mitigated. AMD assumes no obligation to update or otherwise correct or revise this information. However, AMD reserves the right to revise this information and to make changes from time to time to the content hereof without obligation of AMD to notify any person of such revisions or changes.

This information is provided “as is.” AMD makes no representations or warranties with respect to the contents hereof and assumes no responsibility for any inaccuracies, errors, or omissions that may appear in this information. AMD specifically disclaims any implied warranties of non-infringement, merchantability, or fitness for any particular purpose. In no event will AMD be liable to any person for any reliance, direct, indirect, special, or other consequential damages arising from the use of any information contained herein, even if AMD is expressly advised of the possibility of such damages.