Performance Efficiency Today: AMD’s HPC Product Portfolio

Energy efficient CPU and discrete GPU processors focused on addressing the most demanding HPC workloads

**Multi-core x86 Processors**
- Outstanding Performance
- Superior Scalability
- Enhanced Power Efficiency

**ATI FirePro™ Professional Graphics**
- 3D Accelerators For Visualization
- Full support for GPU computation with OpenCL

**AMD FireStream™ GPU Accelerators**
- Optimized for server integration
- Single-slot and dual-slot form factors
- Industry standard OpenCL SDK
### AMD Server Platform Roadmap

<table>
<thead>
<tr>
<th></th>
<th>AMD Opteron™ 6100 Series CPU</th>
<th>&quot;Interlagos&quot; CPU</th>
<th>&quot;Terramar&quot; CPU</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2- and 4-way enterprise, mainstream platform</strong></td>
<td>8/12/16 8/12/16 cores</td>
<td>8/12/16 &quot;Bulldozer&quot; CPU cores</td>
<td>Up to 20 next-generation &quot;Bulldozer&quot; cores</td>
</tr>
<tr>
<td></td>
<td>12M L3</td>
<td>4x HT-3 (6.4GT) DDR3 (quad-channel)</td>
<td>Socket G2012</td>
</tr>
<tr>
<td></td>
<td>4x HT-3 (6.4GT) DDR3 (quad-channel)</td>
<td>4x HT-3 (6.4GT) DDR3 (quad-channel)</td>
<td>Integrated PCIe Gen3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>AMD Opteron™ 4100 Series CPU</th>
<th>&quot;Valencia&quot; CPU</th>
<th>&quot;Sepang&quot; CPU</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1- and 2-way cost-optimized, energy-efficient platform</strong></td>
<td>6/8 &quot;Bulldozer&quot; CPU cores</td>
<td>6/8 &quot;Bulldozer&quot; CPU cores</td>
<td>Up to 10 next-generation &quot;Bulldozer&quot; CPU cores</td>
</tr>
<tr>
<td></td>
<td>4/6-cores</td>
<td>2x HT-3 (6.4GT) DDR3 (dual-channel)</td>
<td>&quot;Bulldozer&quot; CPU cores</td>
</tr>
<tr>
<td></td>
<td>6M L3</td>
<td>2x HT-3 (6.4GT) DDR3 (dual-channel)</td>
<td>Socket G2012</td>
</tr>
</tbody>
</table>

**Timeline:**
- **2010:**
  - "Valencia" CPU
  - "Sepang" CPU
- **2011:**
  - AMD Opteron™ 6100 Series CPU
  - "Interlagos" CPU
- **2012:**
  - AMD Opteron™ 4100 Series CPU
  - "Terramar" CPU

**Process Technologies:**
- 45nm
- 32nm
Power Efficiency on CPUs: AMD-P Technologies

- **AMD PowerCap Manager**: Allows IT datacenter managers to set a fixed limit on a server's processor power consumption.
- **AMD Smart Fetch Technology**: Can reduce power consumption by allowing idle cores to enter a “halt” state.
- **AMD CoolCore™ Technology**: Can reduce processor power consumption by dynamically turning off sections of the processor when inactive.
- **AMD CoolSpeed Technology**: Highly accurate thermal information & thermal protection.
- **Advanced Processor Management Link**: Allows advanced power control and thermal policies.
- **Low Power U/RDDR3 memory**: Supports DDR3 1.5v and low power DDR3L 1.35v memory technologies.
- **Dual Dynamic Power Management**: Enables more granular power management capabilities to reduce processor energy consumption. Separate power planes for cores and memory controller.
- **C1E**: Reduces memory controller and Hypertransport™ technology links’ power.
- **AMD PowerNow!™ Technology with Independent Dynamic Core Technology**: Allows processors and cores to dynamically operate at lower power and frequencies, depending on usage and workload to help reduce TCO and to lower power consumption in the datacenter.
**Designed for Scalability and Performance**

**“Bulldozer” module**
Two cores in a single unit that enables two simultaneous threads, the building blocks of a “Bulldozer” die.

**Parallel Threads**
The ability to execute two threads on two discrete, unshared cores without compromising or increasing bottlenecks.

**Flex FP**
A flexible floating point unit that can be dedicated OR shared between the two cores per cycle.

**Dedicated Scheduler**
Independent integer schedulers and an FP scheduler help improve scalability by efficient execution.
Flex FP: The World’s ONLY Flexible 256-bit FPU

Designed for flexibility and performance in technical computing

**Standard 128-bit mode**
- Two 128-bit FMAC units, one for each core
- Simultaneous execution
- Up to 64 FLOPs per cycle (16-core Interlagos)
- No software recompile
- AVX + FMA4

**Shared AVX mode**
- Single 256-bit AVX unit shared between 2 cores
- Up to 64 FLOPs per cycle (16-core Interlagos)
- AVX + FMA4
AMD FireStream Compute Accelerators

- **Heterogeneous:** Developers leverage AMD GPUs and x86 CPUs for optimal application performance and user experience
- **High performance:** Massively parallel, programmable GPU architecture delivers unprecedented **performance and power efficiency**
- **Industry Standards:** OpenCL™ and DirectCompute enable cross-platform development
**AMD FireStream™ 9370 Delivers 2.64 TFLOPS SPFP**

Single Slot AMD FireStream 9350 Enables Maximum Compute Density

---

### Ultra High-end
- High performance
- Fastest memory technology
- Large memory

### Deployable Performance
- Highest performance per watt and density
- Low power
- Low profile
- Optimal price/

<table>
<thead>
<tr>
<th>Year</th>
<th>Model</th>
<th>Performance</th>
<th>Memory</th>
<th>Power</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>AMD FireStream™ 9250</td>
<td>1.0 TFLOPS</td>
<td>200 GFLOPS DPFP</td>
<td>1GB GDDR3</td>
<td>Single slot, &lt;120W</td>
</tr>
<tr>
<td>2009</td>
<td>AMD FireStream™ 9270</td>
<td>1.2 TFLOPS</td>
<td>240 GFLOPS DPFP</td>
<td>2GB GDDR5</td>
<td>160W</td>
</tr>
<tr>
<td>2010</td>
<td>AMD FireStream™ 9350</td>
<td>2.0 TFLOPS</td>
<td>400 GFLOPS DPFP</td>
<td>2GB GDDR5</td>
<td>Single slot, &lt;150W</td>
</tr>
<tr>
<td>2012</td>
<td>AMD FireStream™ 9370</td>
<td>2.64 TFLOPS</td>
<td>528 GFLOPS DPFP</td>
<td>4GB GDDR5</td>
<td>Passive heat sink</td>
</tr>
</tbody>
</table>

---

*Road map preliminary and subject to change without notice*
The Effect of Ubiquitous High Performance GPU On-die

• The Dilemma for Expanding Traditional GPU Usage Scenarios:
  • Most IGP-based GPU solutions have only supported low-end graphics
  • Higher performance discrete GPUs opened the door for broader usage scenarios
  • But, the attach rates aren’t high enough for the ISV SW community to invest in developing these new scenarios

• Over time, APUs will make high performance GPU attach ubiquitous
  • More applications will make creative use of 3D graphics visualization capabilities
  • Data Parallel Programming languages and tools will enable more developers to use the GPU for more than just 3D graphics
  • The GPU will become more deeply integrated architecturally allowing even more fine-grained computation for new use models
  • As a result, more ISVs will leverage the power efficient data parallel processing and 3D graphics capabilities to create new visually stunning applications

• This kicks a new virtuous cycle into play that is both economically and technically advantageous for new killer applications! This is the promise of Fusion!
Extending the AMD Fusion APU Advantage over Time

Physical Integration
- Integration of CPU & GPU on single silicon chip
- High Bandwidth on-chip Memory Controller
- Open SW Ecosystem to enable Heterogeneous Computing

Optimized Platforms
- Fusion interconnect enhances interface between CPU and GPU
- GPU with high-level language support
- Bi-Directional Power Mgmt CPU/GPU

Architectural Integration
- Unified CPU/GPU Address Space
- GPU uses pageable system memory
- GPU hardware scheduler
- CPU/GPU APU Coherent Memory

Architectural & OS Integration
- GPU compute context switching
- GPU graphics pre-emption
- Coherent PCIe for discrete GPU
- Task Parallel Runtime Integration
Looking Ahead - Ongoing Memory System and GPU Architecture Improvements

- **Significant and ongoing BW improvements throughout the memory system**
- **Continue incredible pace of GPU improvements in performance/watt**
- **Increase sophistication of the GPU so that it becomes a 1st-class citizen of the overall system architecture**
  - Single unified virtual address space
  - Virtual memory support via IOMMU
  - Participation in system-level coherency
  - Support for context switching
Heterogeneous computing software ecosystem

- Based on open industry standards
- Well defined abstractions enable more focused and parallel development
- Encourages aligned contributions from established companies, new companies and universities
- Overarching goal is to enable widespread and highly productive Software development

End-user Applications

High Level Frameworks

Tools: Compilers, Debuggers, Profiles

Domain Libraries
(Math, Video, Imaging, etc.)

OpenCL, DX Runtimes and User Mode Drivers

APU, Discrete CPUs/GPUs & Kernel Drivers
Heterogeneous Computing Software Stack -- *Future*

**End-user Applications**

- **High Level Frameworks**
- **Domain Libraries** (Math, Video, Imaging, etc.)
- **Tools**: Compilers, Debuggers, Profiles

**Task Parallel Queuing Runtimes**
(ConcRT, GCD, TBB etc)

APU, Discrete CPUs/GPUs & Kernel Drivers

Task queuing runtimes replace the driver layer

- GPU is a peer-level programmable processor available to any application
- Abstractions still allow programmers to leverage DSLs to gain higher productivity
Why Fusion Wins in the Long Run

 Superior power efficiency and memory throughput for visual and media applications
 Shared memory architecture is high performance and easy to program
 Quality of Service support for modern Operating Systems
 Applications use cores that match their workloads
 Parallel computation is enabled on the GPU through standards-based SW infrastructure
 Discrete GPUs are FSA compliant and allow further platform-level differentiation

Visit developer.amd.com for more information and Webcast replays from AMD FUSION SUMMIT Developer Conference
DISCLAIMER

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. 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.

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 MERCHANTABILITY OR FITNESS FOR ANY PARTICULAR PURPOSE. IN NO EVENT WILL AMD BE LIABLE TO ANY PERSON FOR ANY 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.

ATTRIBUTION

© 2010 Advanced Micro Devices, Inc. All rights reserved.

AMD, the AMD arrow logo, ATI Radeon, AMD Opteron, and combinations thereof, are trademarks of Advanced Micro Devices, Inc. All other products names and logos are for reference only and may be trademarks of their respective owners.

Third-party products and services listed for illustrative purposes only. No endorsement is implied.