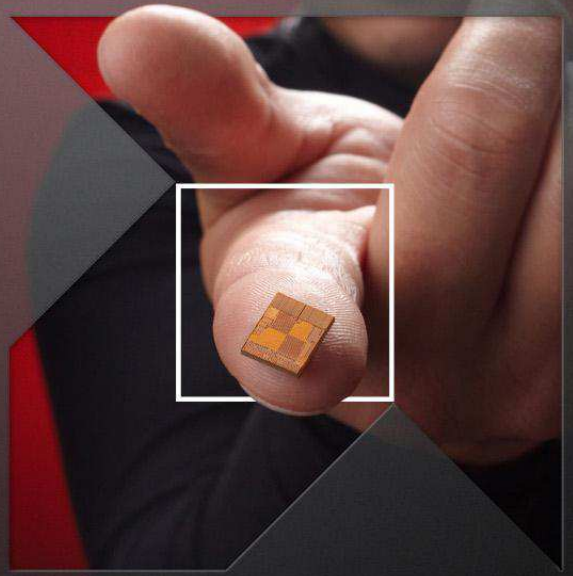


THE SURROUND COMPUTING ERA

Mark Papermaster
Senior Vice President and CTO, AMD

Hot Chips Symposium
Cupertino, CA
August 28, 2012

AMD

A RAPIDLY CHANGING ENVIRONMENT

Users want content anytime, any platform, anywhere

Explosion of unstructured data

- 245 exabytes of data crossed Internet in 2010[1]
- Growing to 1000 exabytes in 2015

Data center server demand >10M units by 2016[2]

1. Cisco Visual Networking Index Global IP Traffic Forecast, 2010 to 2015
2. Worldwide and Regional Server 2012-2016 Forecast, IDC, May 2012

AMD

REVOLUTIONARY TRANSFORMATION

10 years ago: **The Interactive Computing Revolution**

- Graphics acceleration enabled
- Computing accessible to everyone
- Touch screen phones to cinematic 3D

Starting now: **The Surround Computing Era**

- Computers are everywhere
- Integrating into our environment
- Computing is part of everyday life, not a distinct activity

AMD

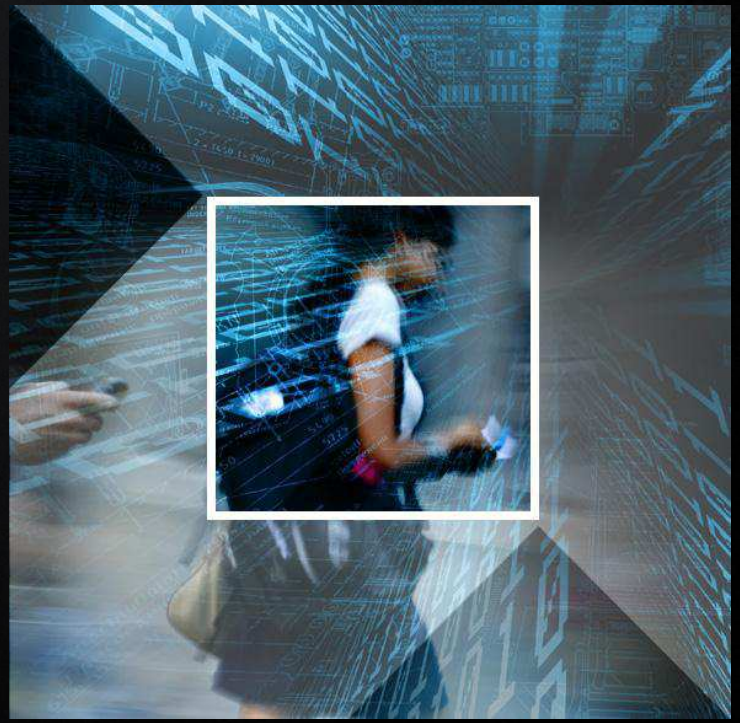
SURROUND COMPUTING

We are entering the Surround Computing Era

- Multi-platform – eyeglasses to room-size
- Fluid – realistic output, natural human input
- Intelligent – anticipates our needs

Profound implications for computer architecture

- Smarter clients – realistic, natural human communication
- Smarter clouds – orchestrate 10B devices in real-time



AMD

SMARTER CLIENTS

Natural UI and Gestures

Touch, gesture and voice

Biometric Recognition

Secure, fast, accurate: face, voice, fingerprints

Augmented Reality

Superimpose graphics, audio, and other digital information as a virtual overlay

Content Everywhere

Content from any source to any display seamlessly

Beyond HD Experiences

Streaming media, new codecs, 3D, transcode, audio

AV Content Management

Searching, indexing and tagging of video and audio. Multimedia data mining

New Surround Compute Applications and Experiences – Accelerators Required!

AMD

SMARTER CLOUDS

The Cloud is
the "Backbone" of
Surround Computing

Surround Computing Cloud Services

Trust

Context

Analytic
Compute

Consistent Experiences Across Multiple Devices

AMD

SMARTER CLOUDS – MANAGING UNSTRUCTURED DATA

Connected devices drive cloud computational loads

Datacenter optimized for Surround Computing

- **Scale** – to support tens of billions of connected devices
- **Acceleration** – back-end NUI, graphics, analytics
- **Security, privacy** – consistent end-to-end architecture
- **Real time** – latency is critical
- **Dense servers** – optimized for low power

AMD

THE WAY FORWARD

Surround computing

- Requires smarter clients and clouds
- Efficient datacenters

Heterogeneous engines

- Accelerate key client and server parallel workloads

Heterogeneous System Architecture (HSA)

- New silicon architecture making it all work together

AMD

CHANGING THE THINKING, CHANGING THE GAME

HSA – directly access acceleration hardware

- Unlocks the value of the GPU to software developers
- Program in C, C++, Java, Python, JavaScript, HTML5
- ISA agnostic

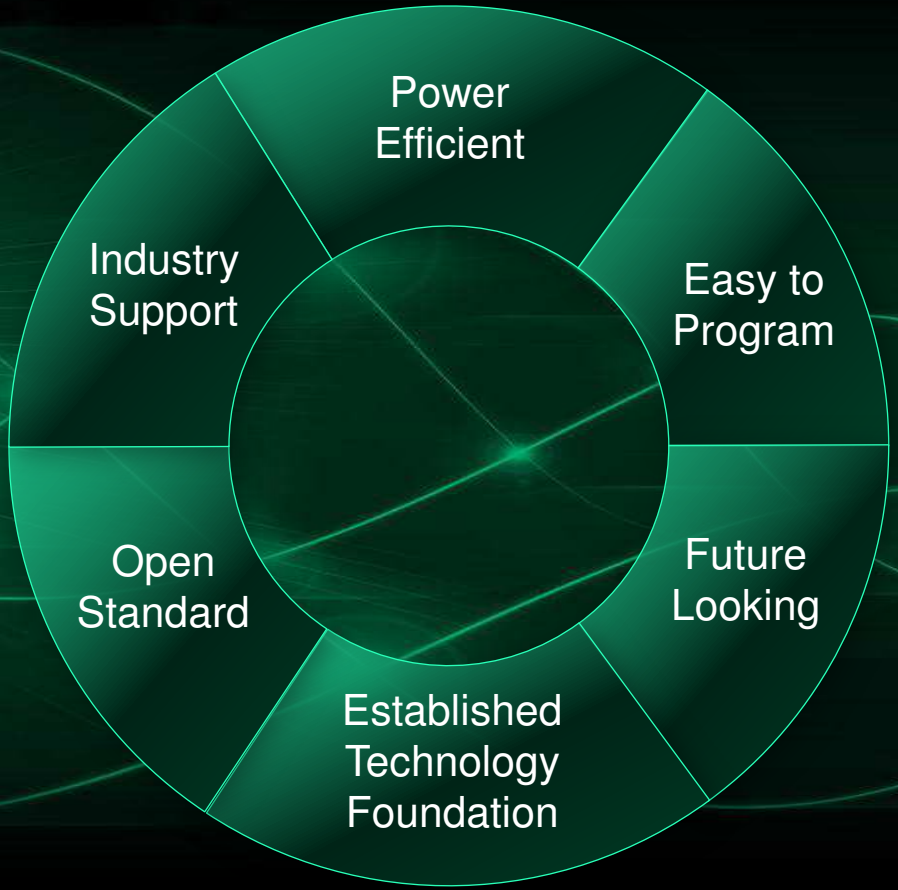
GPU = CPU in terms of processing capability

- Full programming language features
- Shared virtual memory: pointer is a pointer
- Coherency and context switching

HSA Foundation is an industry-wide initiative

AMD

BENEFITS OF HETEROGENEOUS SYSTEM ARCHITECTURE



AMD

HSA MEANS ACCELERATED PROCESSING UNITS (APU)

APU is the breakthrough app enabler

APU enables parallel compute and HSA

Emerging workloads require:

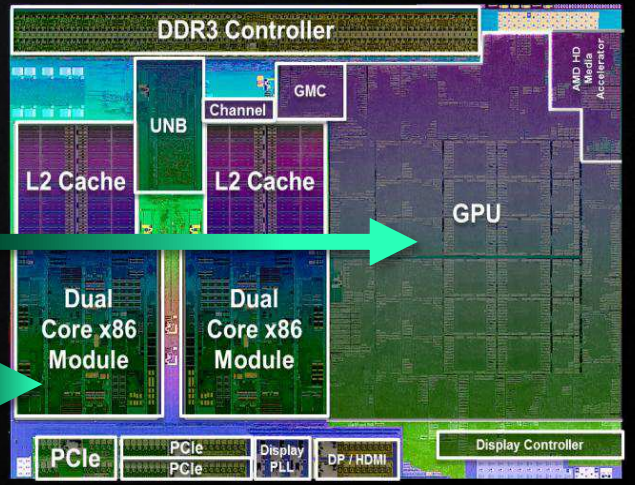
- **Seamless execution across CPU/GPU**
- **Other specialized engines**

APU is the platform of choice

Data Parallel Workloads

Serial and Task Parallel Workloads

HSA Accelerated Processing Unit



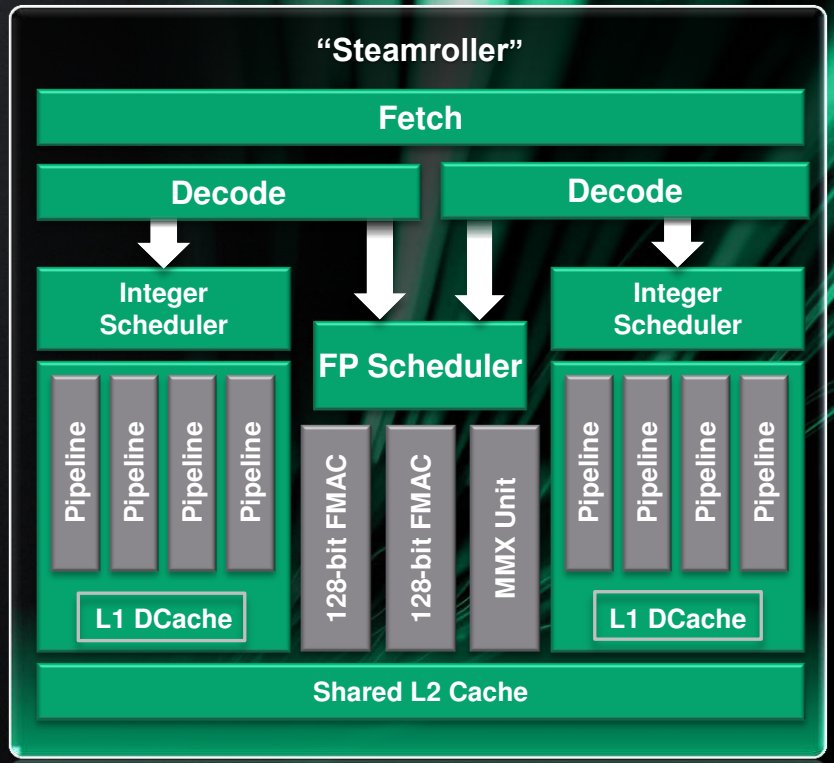
AMD

AMD "STEAMROLLER" CORE

Multi-threaded microarchitecture

Expands computation efficiency

- Feed the cores faster
- Improve single-core execution
- Push on performance/watt

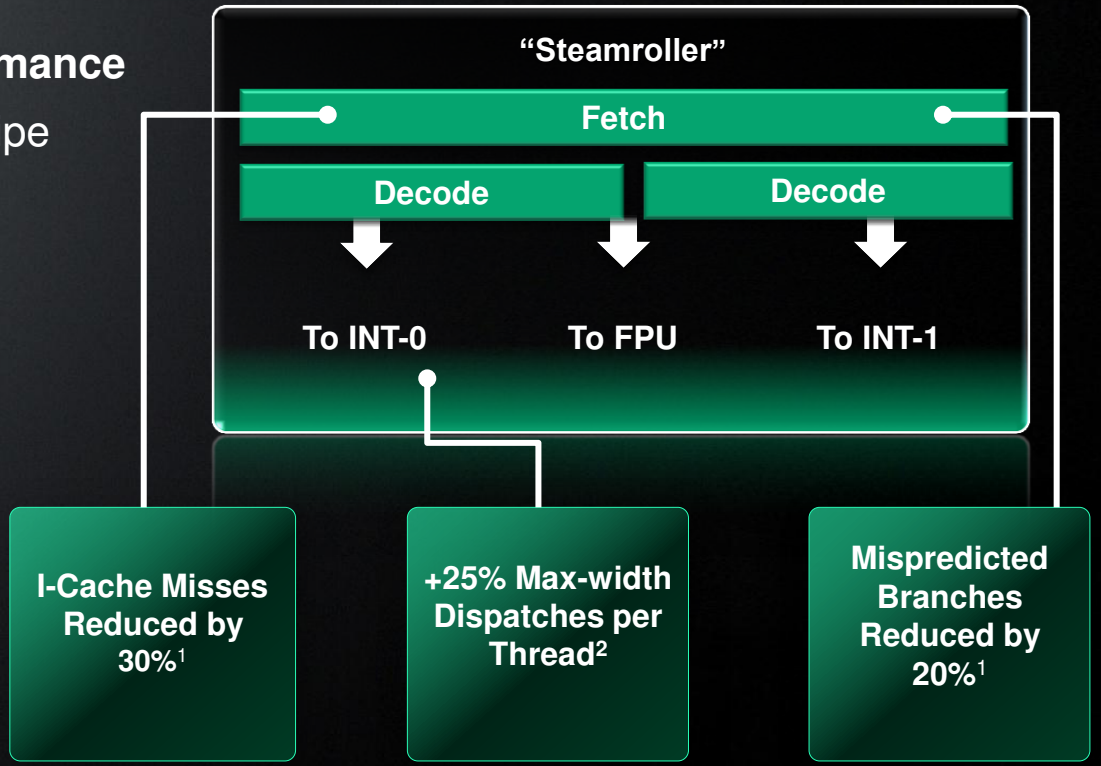


"STEAMROLLER": FEED THE CORES FASTER

No compromises two thread performance

- Dedicated decode for each integer pipe
- Increase instruction cache size
- More efficient dispatch
- Enhance instruction pre-fetch

30% Ops per Cycle Improvement[2]



1.Based on AMD's internal simulation results of average workloads of simulated performance on a number of tests, including those testing transaction processing. (Systems have to be publicly available to publish SPEC CPU Rate.)
2.Based on AMD's internal simulation results of average workloads of simulated performance on a number of tests, including those digital media, productivity and gaming applications.

AMD

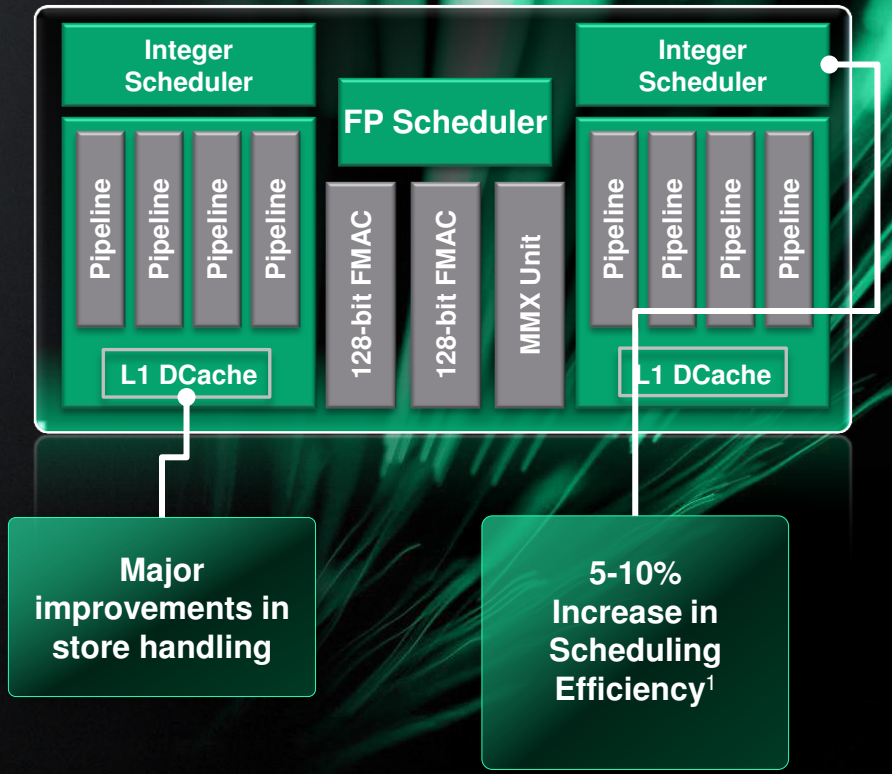
"STEAMROLLER": IMPROVING SINGLE-CORE EXECUTION

Design to tune up integer execution bandwidth:

- **In concert with feeding the core faster**
- More register resources, same latency
- More intelligent scheduling

Design to decrease average load latency:

- **Minimum latency is only part of story**
- Faster handling of data cache misses
- Accelerate store-to-load forwarding



AMD

"STEAMROLLER" PERFORMANCE/WATT DESIGN

Microarchitectural power optimization

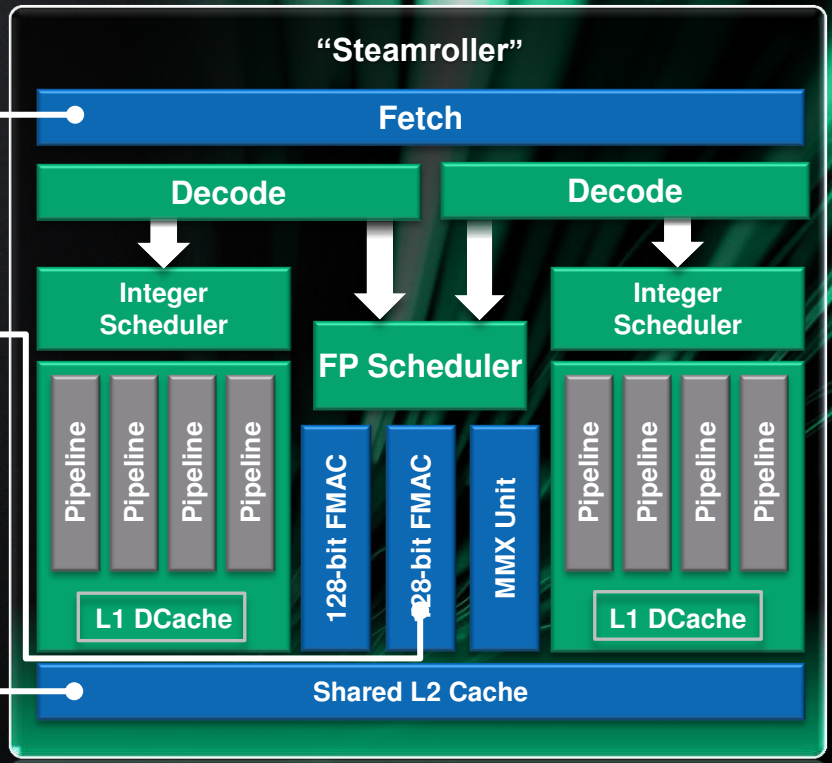
- Lower average dynamic power
- Optimize for loop behaviors

Floating point rebalance

- Streamlined execution hardware
- Adjust to application trends

Dynamic resizing of L2 cache

- Adaptive mode based on workload



AMD

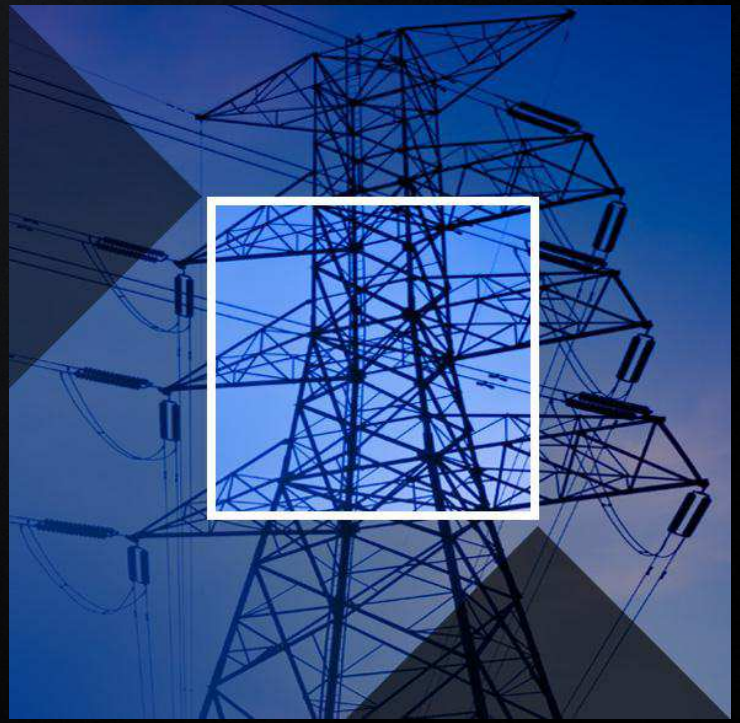
SMART DESIGNING FOR LOW POWER

Power efficiency is fundamental

- Long battery life
- Sleek, light weight form factors
- Cool and quiet computation
- Lower energy consumption and utility bills
- Lower data center TCO

Multi-faceted attack beyond process technology

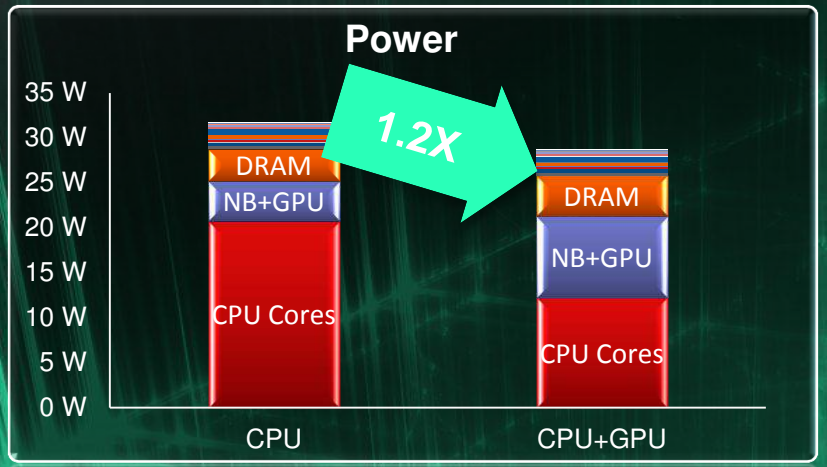
- Optimize hardware with software applications
- Intelligent on-die power management
- Efficient design methodologies

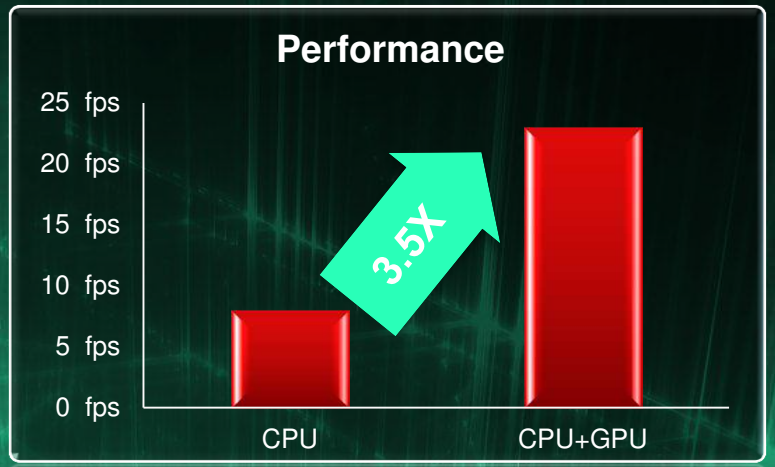


AMD

ARCHITECTURAL EFFICIENCY EXAMPLE WITH VIDEO ENHANCEMENT

MOTION DSP 720P





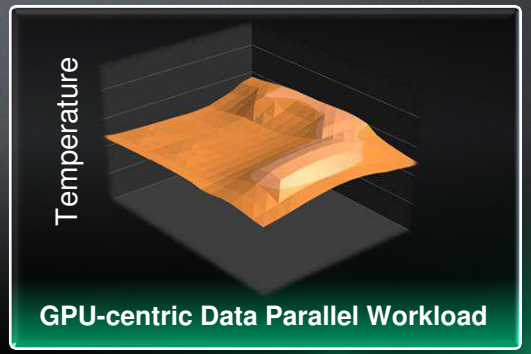
Synergistic use of GPU compute + shared memory = lower power *and* higher performance

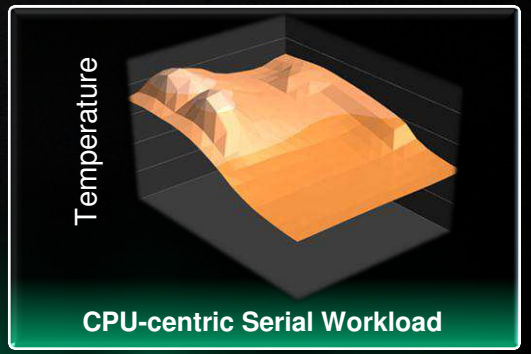
>4.0X Better Energy Efficiency[1]

1. AMD ES-3200 APU)Llano-32nm, 2 cores @ 2400Mhz, GPU:2 CU @ 444Mhz), Windows 7 OS, MotionDSP vReveal Applications (http://www.vreveal.com/stabilization)

POWER MANAGEMENT: FINE GRAINED DYNAMIC POWER TRANSFERS

AMD incorporates activity-based power transfer between CPU and GPU

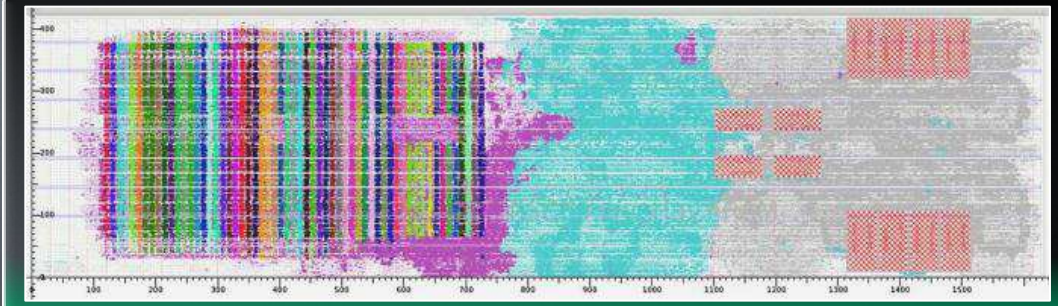




Enabled by sophisticated on-die microcontroller and sensors

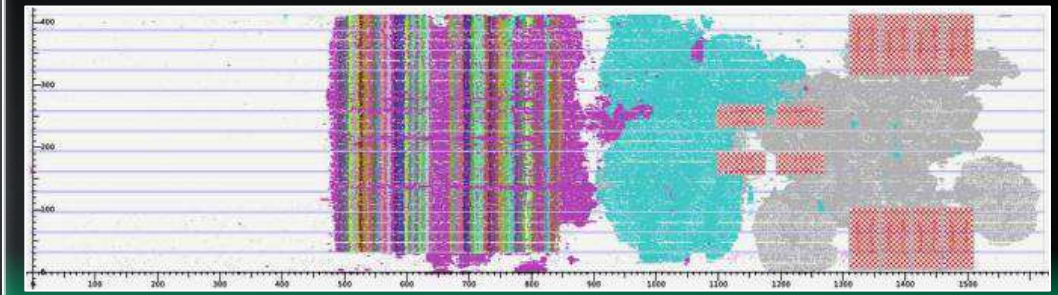
AMD

POWER EFFICIENCY GAINS FROM IMPROVED DESIGN METHODS



“Bulldozer”

Part of the Floating Point Unit. Hand-drawn for maximum speed and density in 32nm



With High Density Library

The same blocks again, but rebuilt using a **High-Density** cell library to achieve **30% area and power reductions**

15-30% lower energy per operation[1] for power constrained designs – same order as a full process node improvement

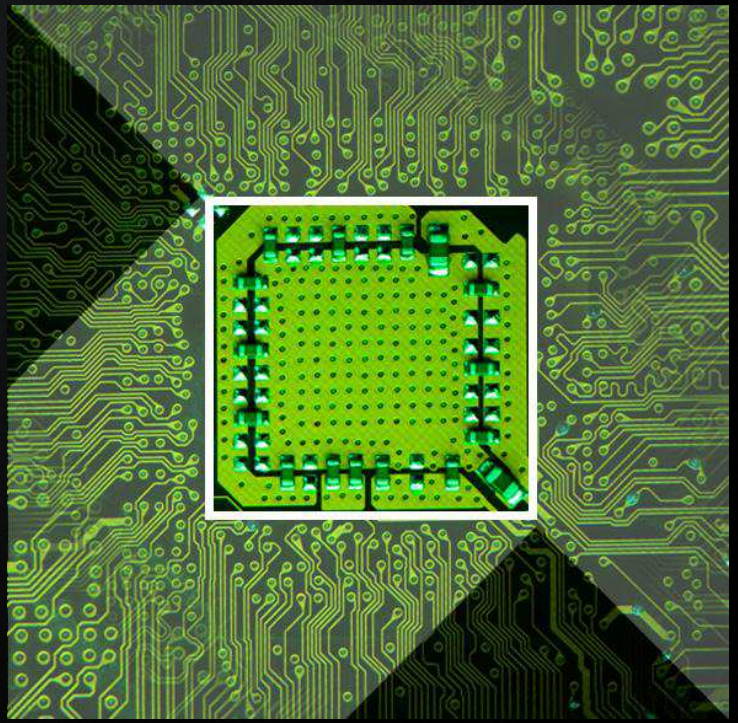
1. Based on AMD’s internal analysis

AMD

FAST FABRICS TIE EVERYTHING TOGETHER

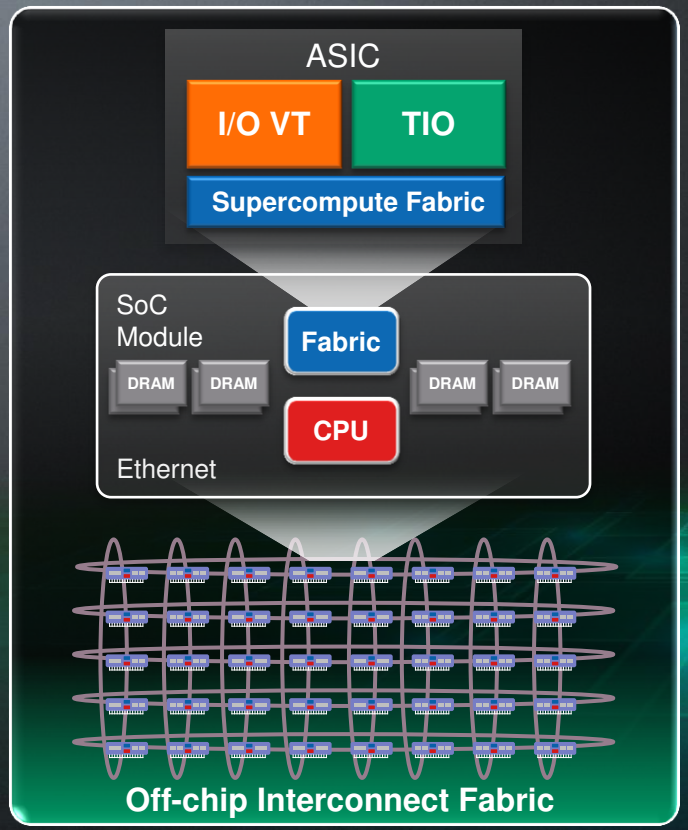
Great interconnect fabrics are needed

- Optimally process unstructured data
- Able to connect massive numbers of processors
- Lowest possible overhead



AMD

AMD FREEDOM FABRIC™ TECHNOLOGY

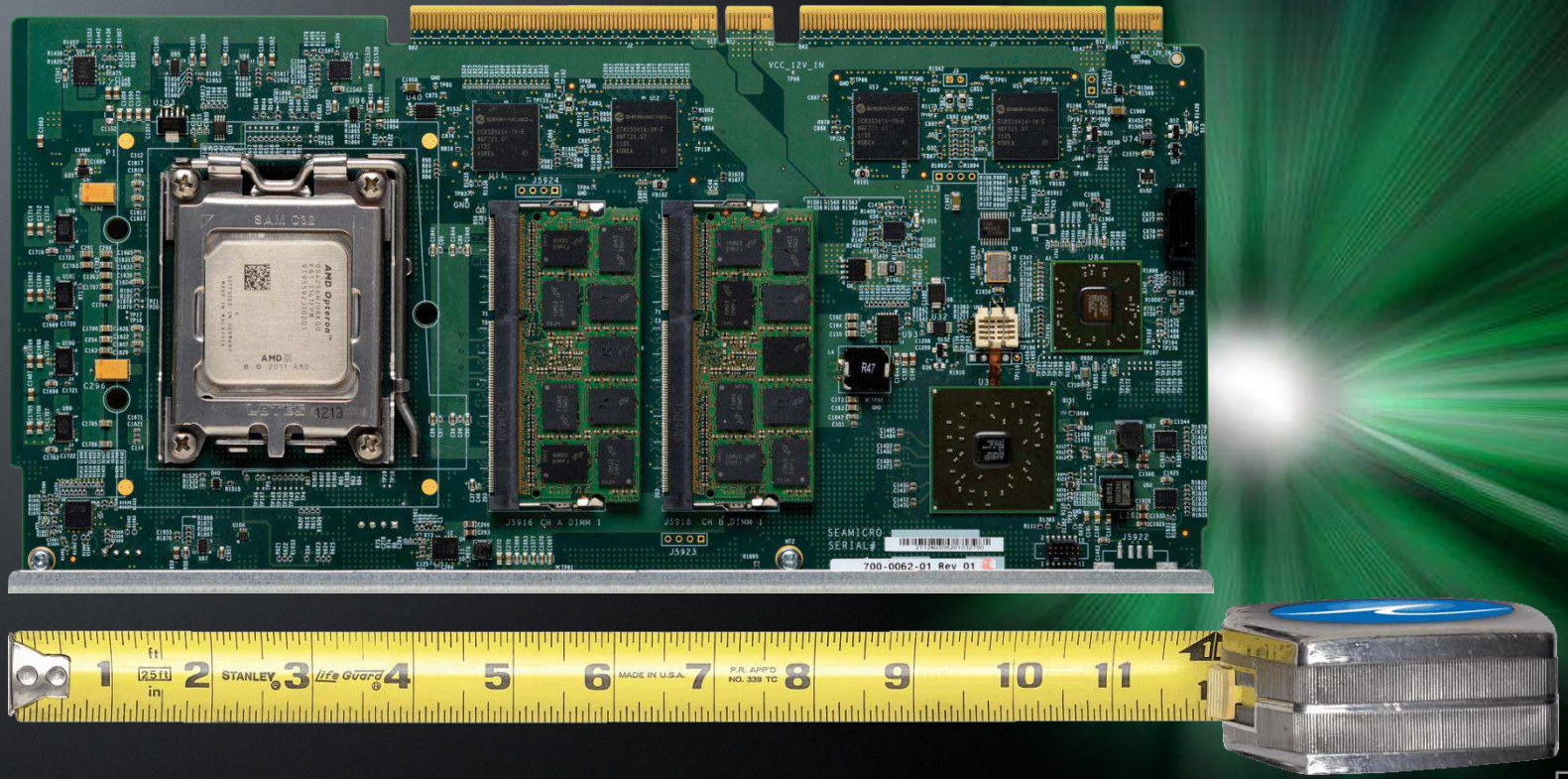


AMD off–chip interconnect fabric IP

- Designed to enable significantly lower TCO
- Links hundreds ➡ thousands of SoC modules
- Shares hundreds of TBs storage and virtualizes I/O
- 160Gbps Ethernet Uplink
- Instructions Set Architecture agnostic

AMD

END-TO-END SYSTEM OPTIMIZATION



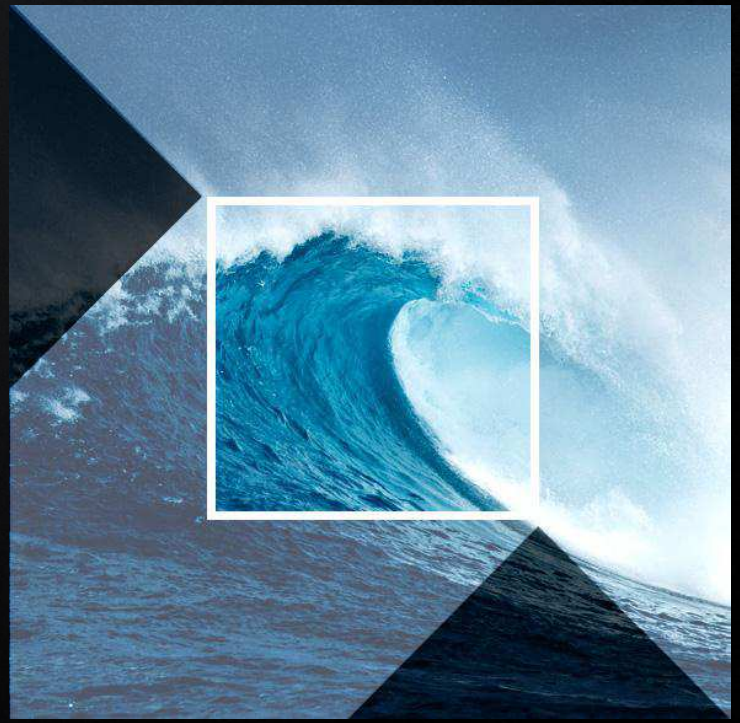
THE PURE SPEEDS AND FEEDS RACE IS OVER – IT'S ABOUT THE SOLUTION!

- End-to-end system view
- Acceleration of the application stack
- Agile delivery of tailored solutions
- Leveraging differentiated IP

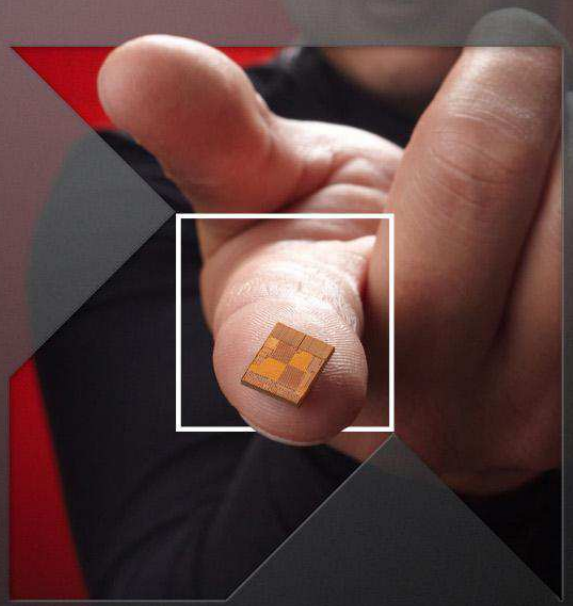
AMD

THE NEXT WAVE – SURROUND COMPUTING REVOLUTION

- AMD products will enable the transition
  - HSA
  - Ambidextrous
  - Fast fabrics
  - Relentless focus on power efficiency
- AMD inspired the interactive computing revolution
- Now leading the way to surround computing



AMD



THANK YOU

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

© 2012 Advanced Micro Devices, Inc. All rights reserved. AMD, the AMD Arrow logo, Radeon, and combinations thereof are trademarks of Advanced Micro Devices, Inc. Other names and logos are used for informational purposes only and may be trademarks of their respective owners.