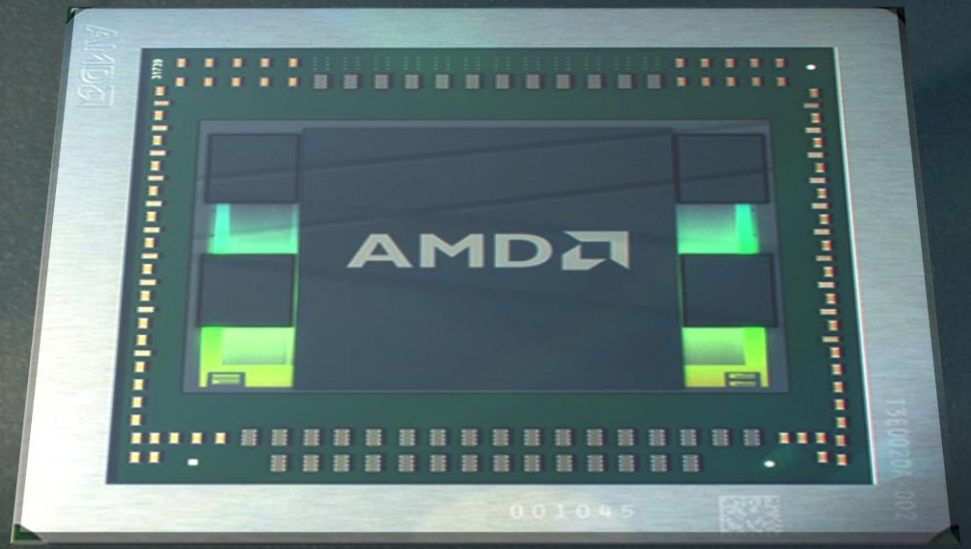
AMD's Next Generation GPU and High Bandwidth Memory Architecture: FURY

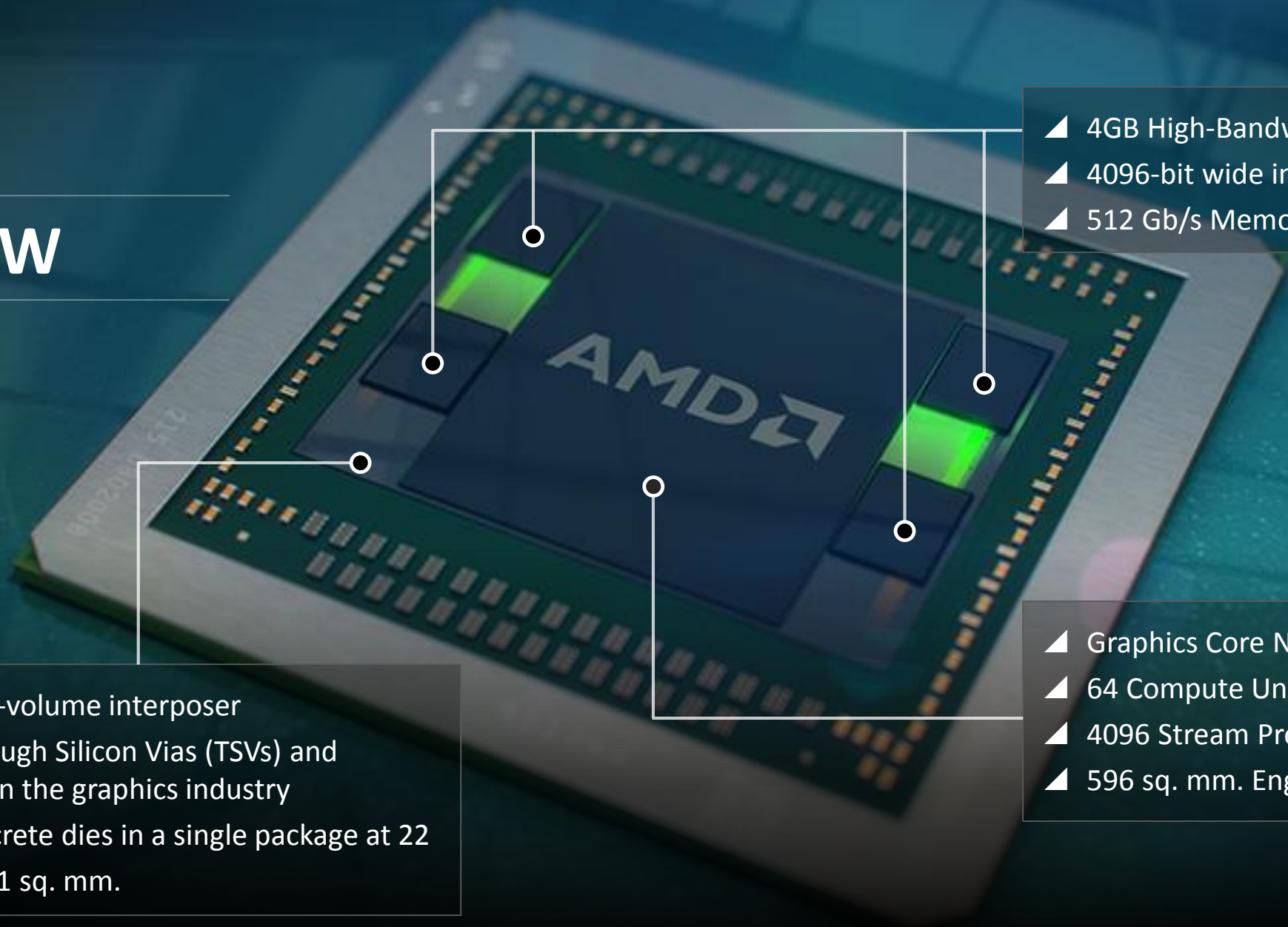


Joe Macri, AMD Corporate Fellow and
Chief Technology Officer, AMD Products

AGENDA

- Overview of AMD Radeon R9 Fury
- Why HBM and Die Stacking
- The Journey to Fury
- Performance
- Form Factor Innovation

OVERVIEW



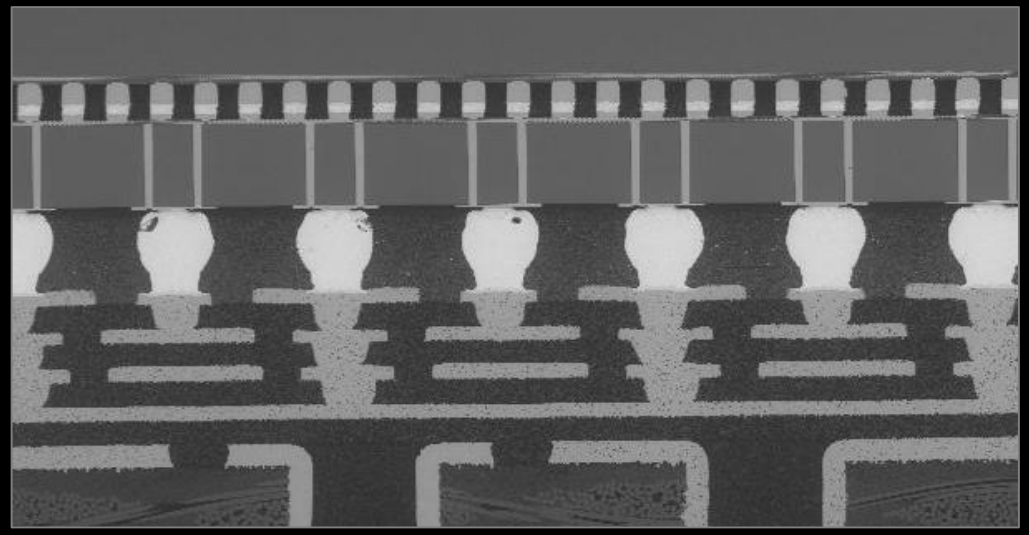
- 4GB High-Bandwidth Memory
- 4096-bit wide interface
- 512 Gb/s Memory Bandwidth

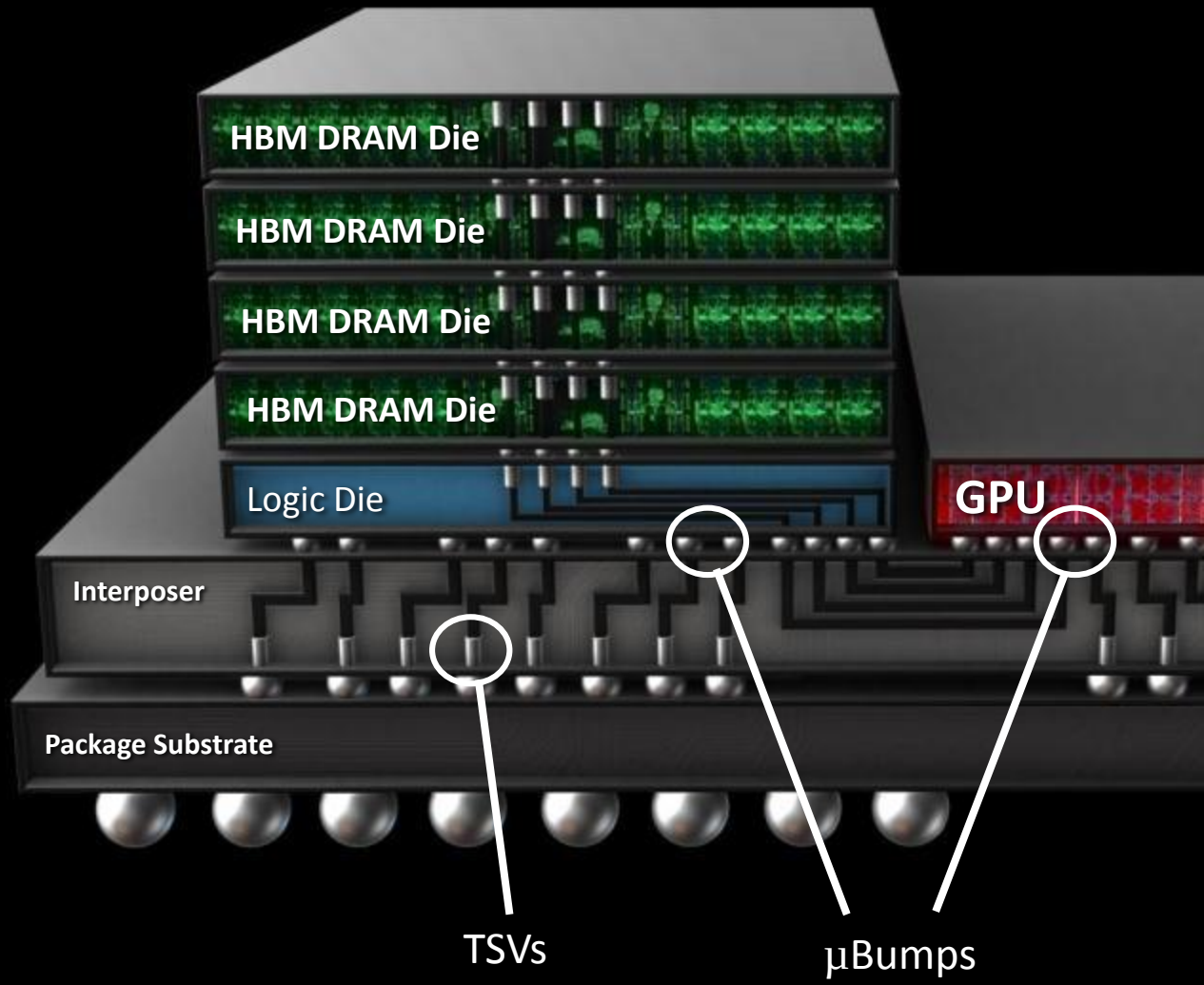
- First high-volume interposer
- First Through Silicon Vias (TSVs) and µBumps in the graphics industry
- Most discrete dies in a single package at 22
- Total 1011 sq. mm.

- Graphics Core Next Architecture
- 64 Compute Units[1]
- 4096 Stream Processors
- 596 sq. mm. Engine

DIE STACKING TECHNOLOGY

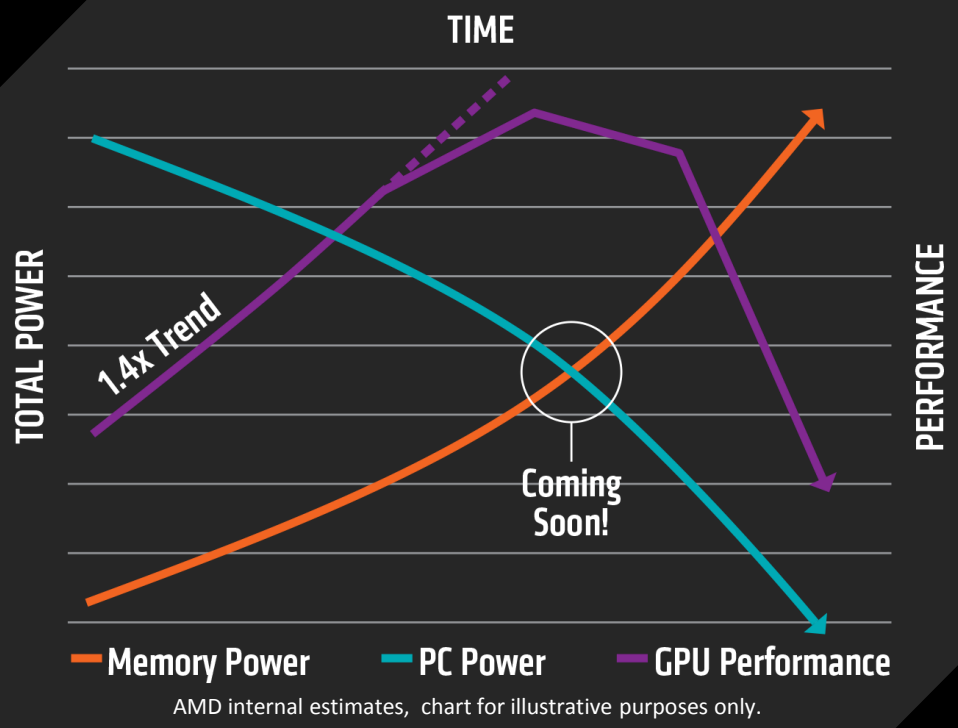
- Die stacking facilitates the integration of discrete dies
- 8.5 years of technology development at AMD and its partners





THE GRAPH THAT STARTED IT ALL

- Platforms & devices must balance power usage between DRAM and logic chips
- GDDR5 is entering the inefficient region of the power/performance curve
- AMD anticipated this challenge seven years ago and began work on a solution

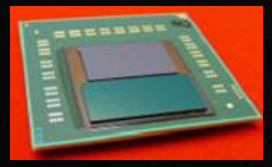


AMD internal estimates, chart for illustrative purposes only.

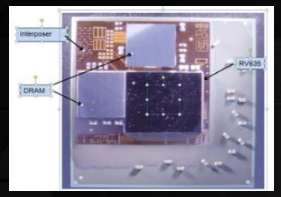
PROTOTYPING ACTIVITIES OVER TIME

AMD

First Time Out



CPU + D3 Mech.



dGPU + G3

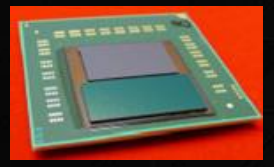
Component reliability: TC | uHAST | HTS

2007 **(100's of samples)** **2011** **(<5000 of samples)** **Jul'14** **(>5000 samples)** **2014**

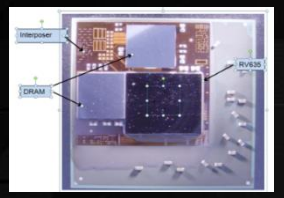
PROTOTYPING ACTIVITIES OVER TIME

AMD

First Time Out



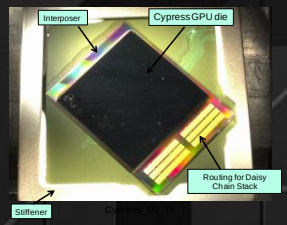
CPU + D3 Mech.



dGPU + G3

Primary Learning

345mm² ASIC
500mm² IP



dGPU

ESD | BLRT | Sort

PwrCyc | uBump EM | TSV EM/SM

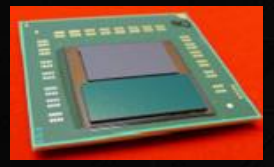
Component reliability: TC | uHAST | HTS

2007 (100's of samples) **2011** (<5000 of samples) **Jul'14** (>5000 samples) **2014**

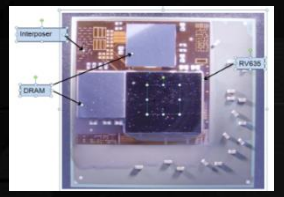
PROTOTYPING ACTIVITIES OVER TIME

AMD

First Time Out



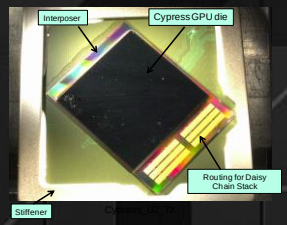
CPU + D3 Mech.



dGPU + G3

Primary Learning

345mm² ASIC
500mm² IP



dGPU

Began migration to larger dGPUs

- 502mm² ASIC w/ 818mm² interposer
- Larger than reticle interposer
- CPI of large system
- Finalize TMP details

ESD | BLRT | Sort

Cost Down

PwrCyc | uBump EM | TSV EM/SM

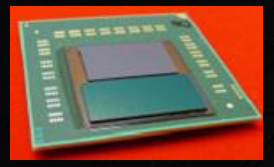
Component reliability: TC | uHAST | HTS

2007 **(100's of samples)** **2011** **(<5000 of samples)** **Jul'14** **(>5000 samples)** **2014**

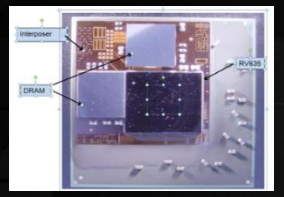
PROTOTYPING ACTIVITIES OVER TIME

AMD

First Time Out



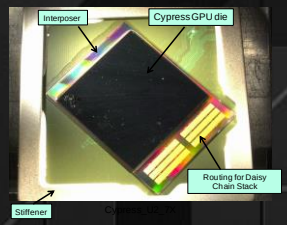
CPU + D3 Mech.



dGPU + G3

Primary Learning

345mm² ASIC
500mm² IP

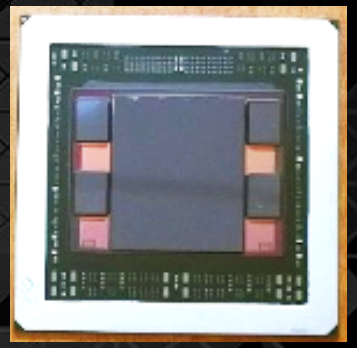


dGPU

502mm² ASIC
818mm² IP

dGPU

Product Readiness



Fiji Replica

592mm² ASIC
1011mm² IP

ESD | BLRT | Sort

Cost Down

PwrCyc | uBump EM | TSV EM/SM

Component reliability: TC | uHAST | HTS

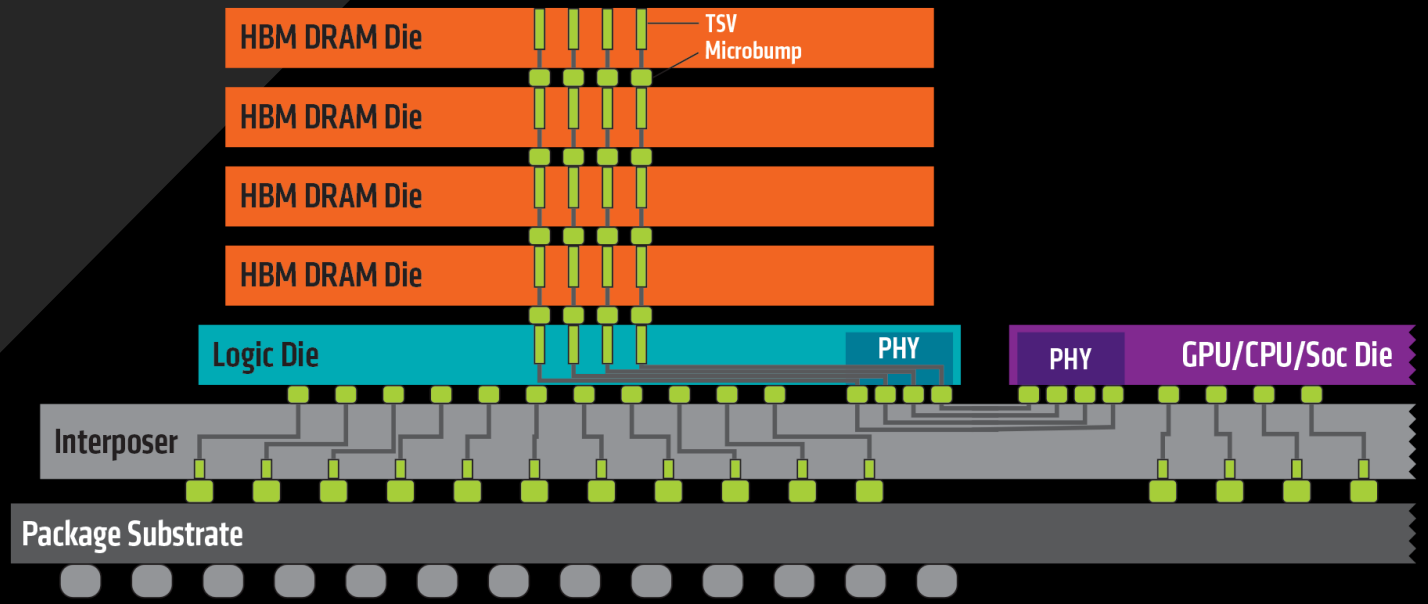
2007 **(100's of samples)** **2011** **(<5000 of samples)** **Jul'14** **(>5000 samples)** **2014**

HIGH-BANDWIDTH MEMORY

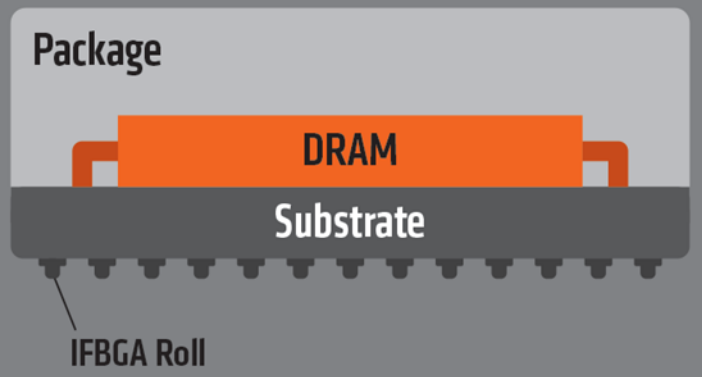
DRAM BUILT FOR AN INTERPOSER

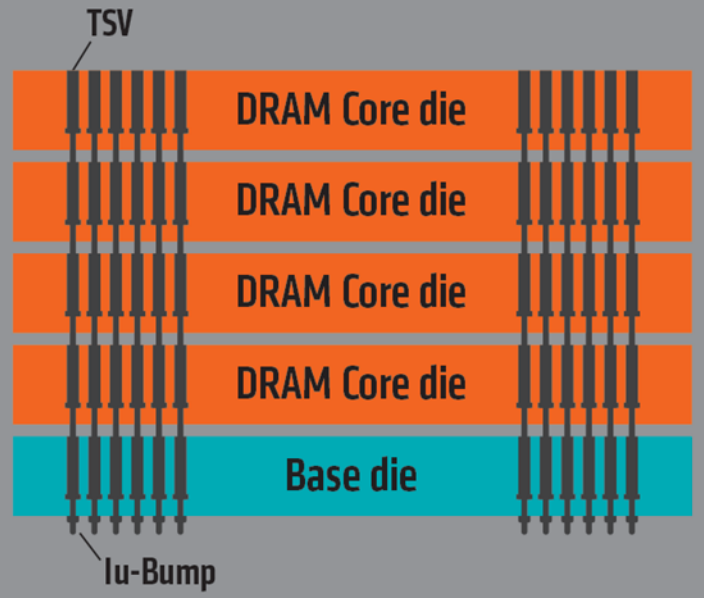
AMD

- A new type of memory chip with low power consumption and an ultra-wide bus width
- Many of those chips stacked vertically like floors in a skyscraper
- New interconnects, called “through-silicon vias” (TSVs) and “µbumps”, connect one DRAM chip to the next
- TSVs and µbumps also used to connect the SoC/GPU to the interposer
- AMD and SK Hynix partnered to define and develop the first complete specification and prototype for HBM



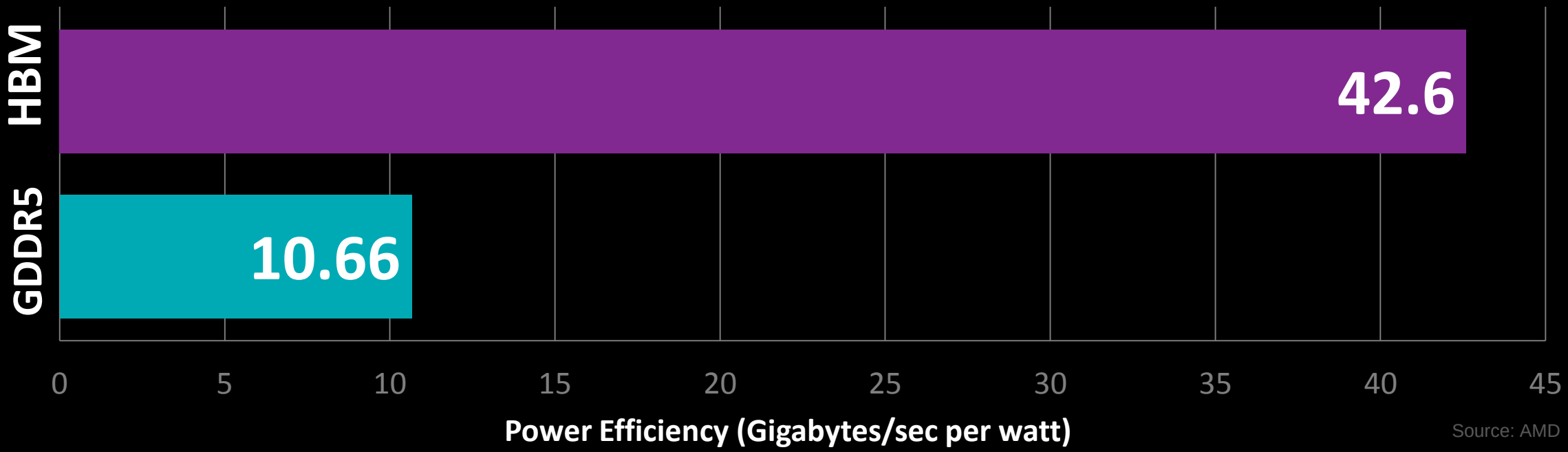
HBM: A DIFFERENT MEMORY FROM GDDR5





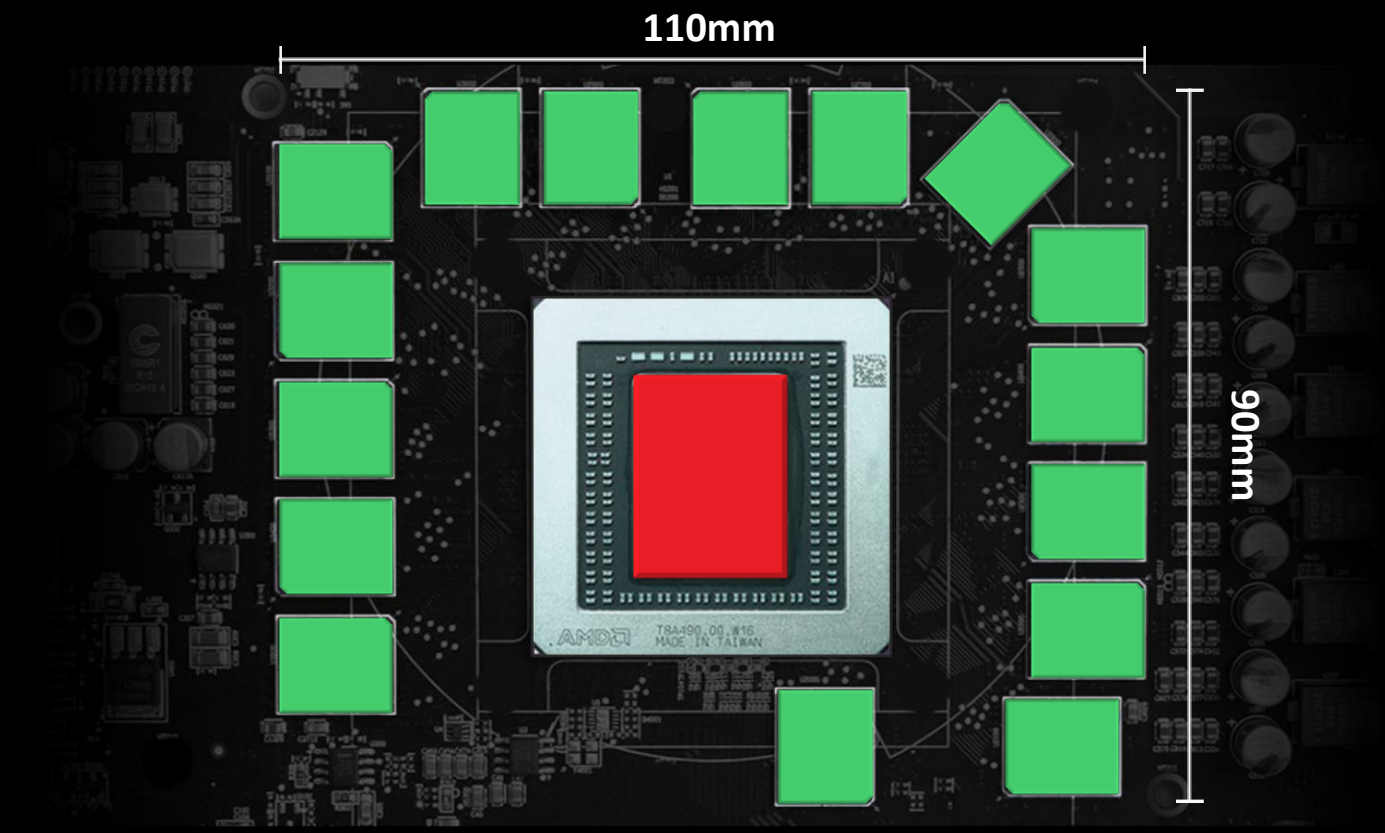
GDDR5	Per Package	HBM
32-bit	Bus Width	1024-bit
Up to 1750MHz (7GBps)	Clock Speed	Up to 500MHz (1GBps)
Up to 28GB/s per chip	Bandwidth	>100GB/s per stack
1.5V	Voltage	1.3V

IMPROVING POWER EFFICIENCY WITH STACKED HBM

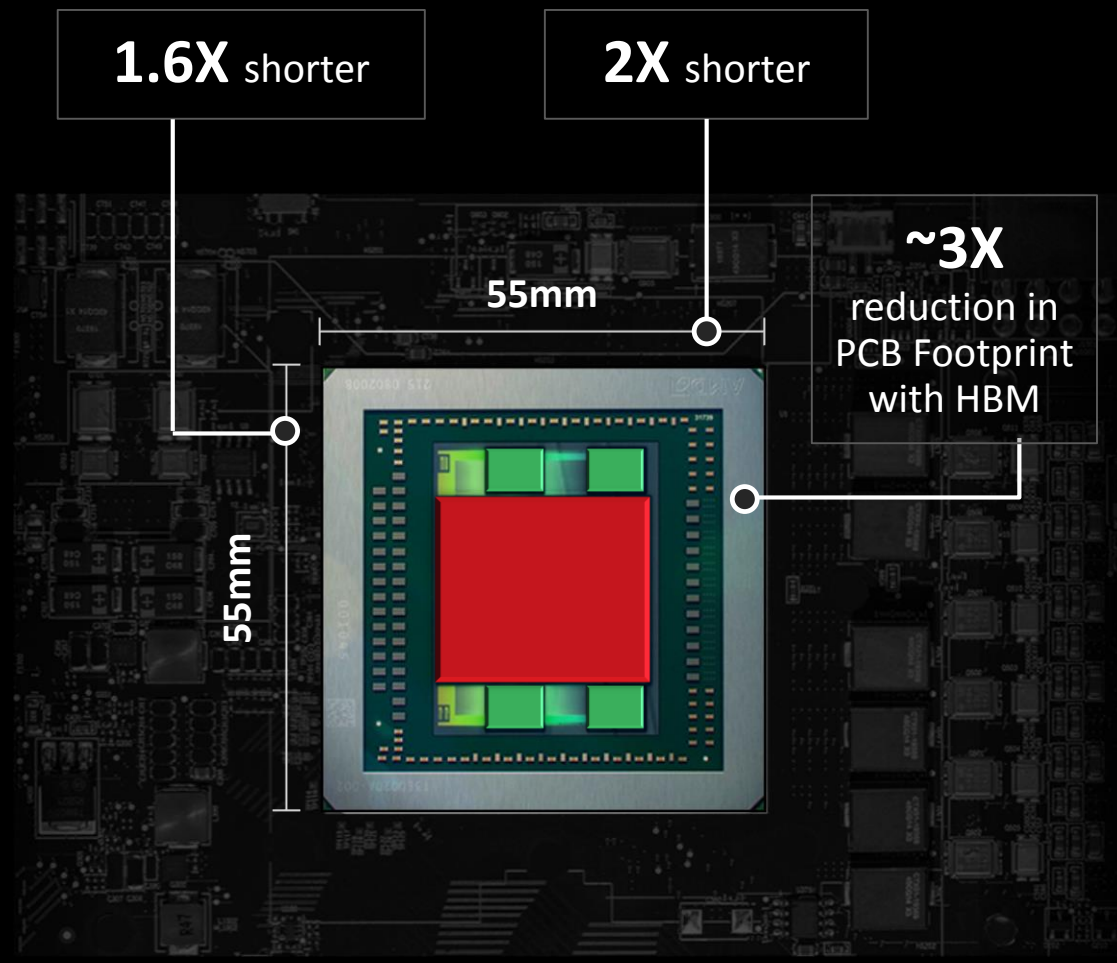


- HBM & interposer give 60% more bandwidth than GDDR5 for 60% less power[2]
- HBM rebalances DRAM vs. logic power consumption to protect future GPU performance growth

EFFICIENT DESIGN



PCB area occupied by ASIC + Memory (Radeon™ R9 290X)



PCB area occupied by ASIC with HBM

RADEON R9 FURY

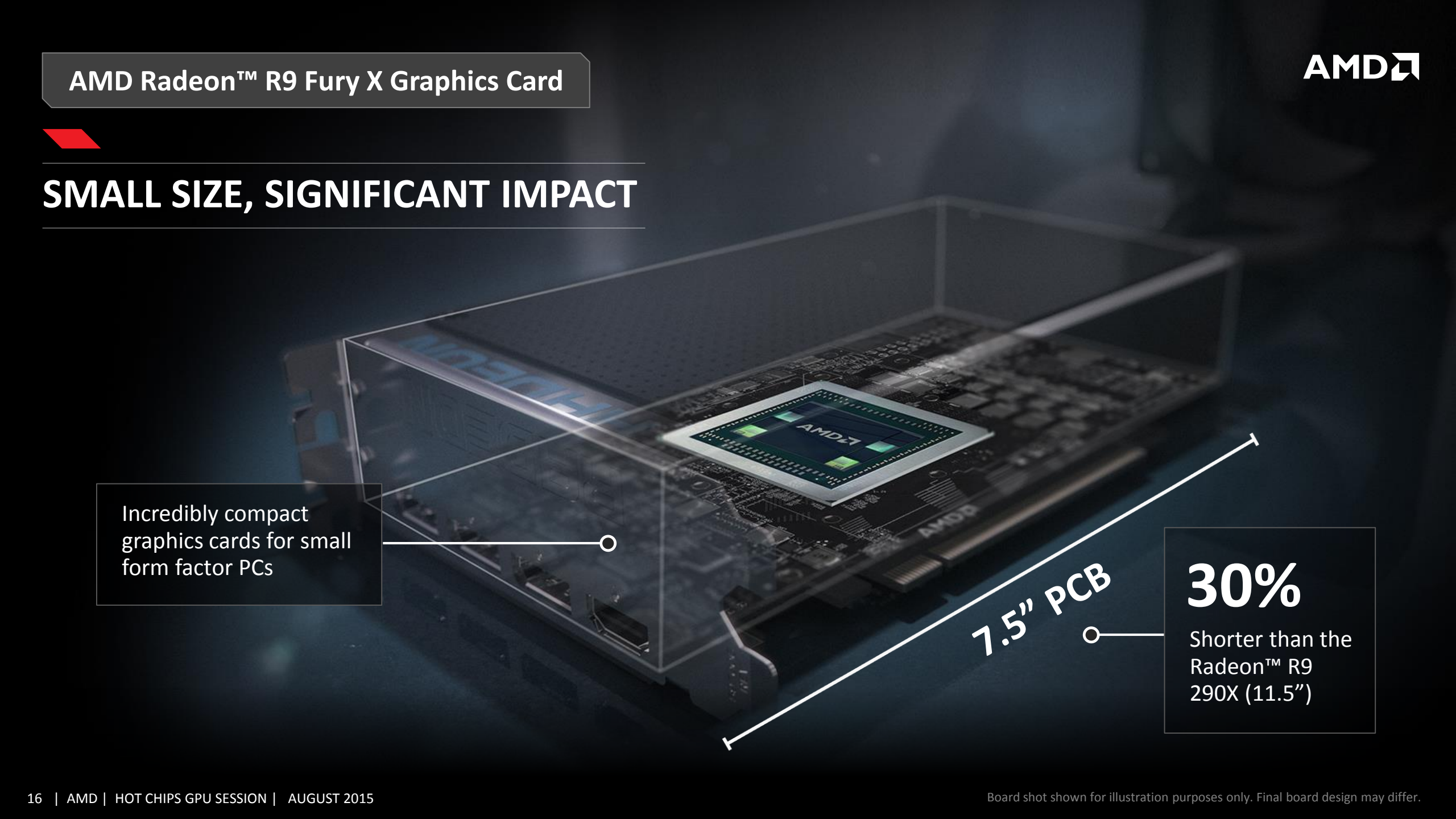
- Configured for Gaming
  - 4 Prims/cycle
  - 64 Pix/cycle
  - 64CU
  - 4096 Ops/Cycle
    - 1/16 rate DPFP
  - 256 Texture/cycle
  - 2MB L2
  - Compute wave switch
- HBM
- Delta Color Compression
- Tessellation Redistribution
- SRIOV Virtualization
- 4 core Async Compute
- Dispatch Draw
- TSMC 28nm HPX
- Improved clock gating
- 1.5x Perf/Watt over Hawaii

GENERATIONAL COMPARISONS

AMD

	Radeon R9 Fury X	Radeon R9-290X	Increase
SPFP Compute	> 8 TFLOPS	5.6 TFLOPS	> 1.4x
Peak Bandwidth	512 GB/sec	320 GB/sec	1.6x
Texture fill rate	> 256 Gtexels/sec	176 Gtexels/sec	> 1.45x
Die area	593 mm^2	438 mm^2	1.35x
Peak SP GFLOPS/mm^2	> 13.5	12.8	> 1.05x
L2 Cache Capacity	2 MB	1 MB	2x

SMALL SIZE, SIGNIFICANT IMPACT



Board shot shown for illustration purposes only. Final board design may differ.

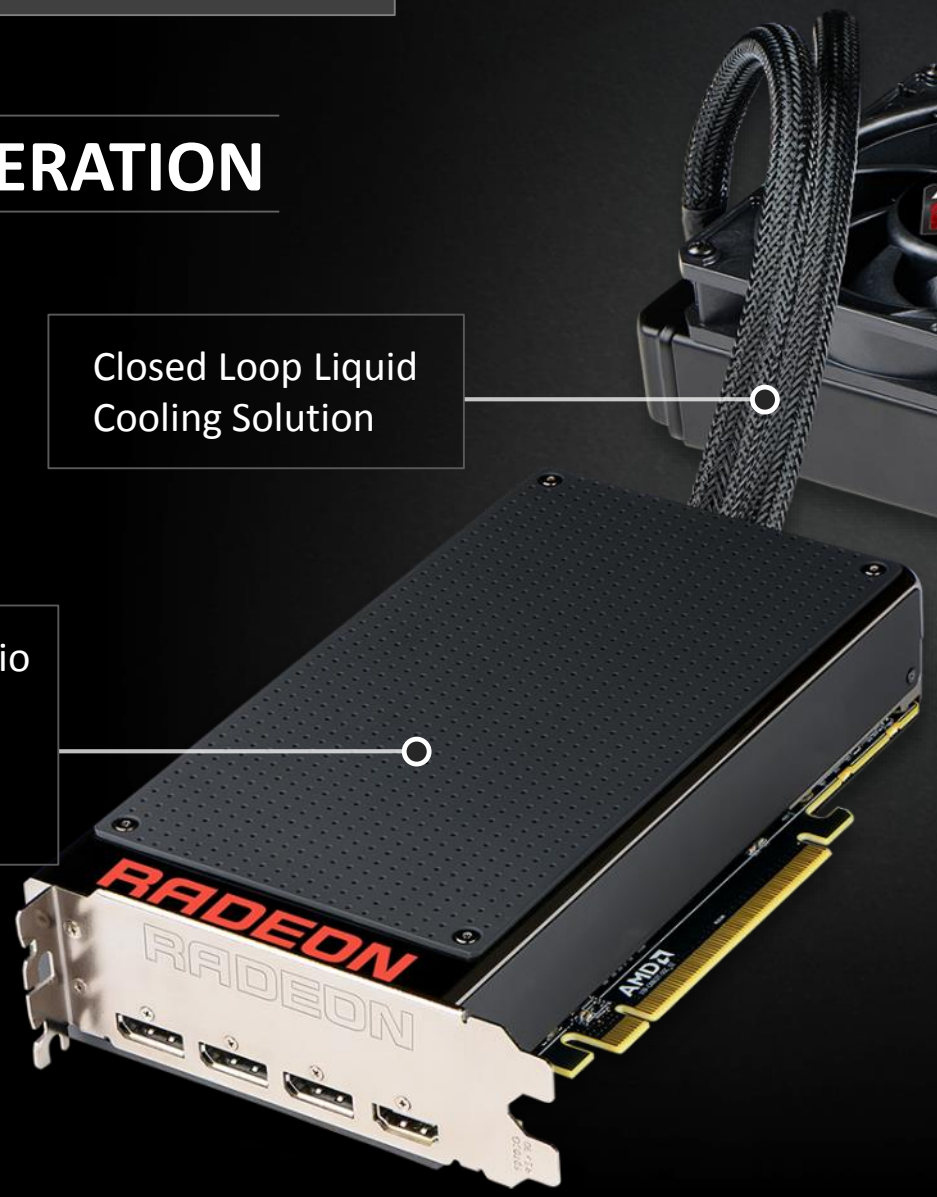
AMD Radeon™ R9 Fury X Graphics Card

COOL & QUIET OPERATION

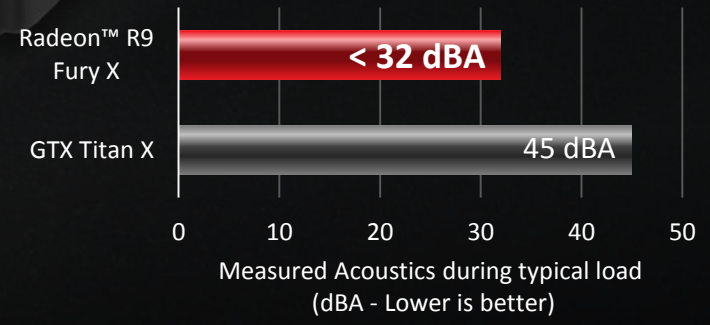
Closed Loop Liquid Cooling Solution

Typical gaming scenario GPU temperature

~50°C



High quality 120mm fan

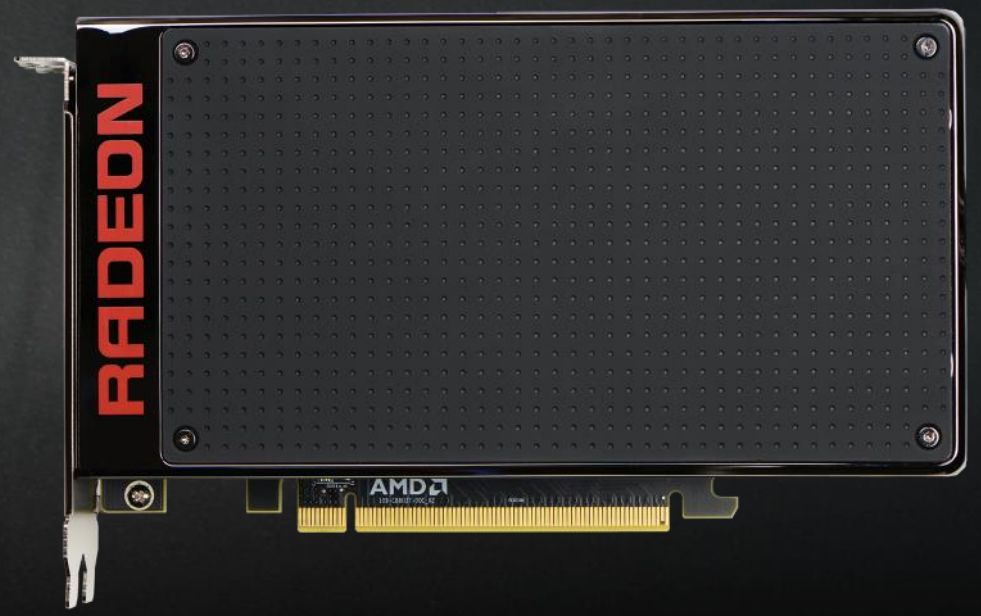


Board shot shown for illustration purposes only. Final board design may differ.

AMD Radeon™ R9 Fury X Graphics Card

4K ULTRA SETTINGS SMOOTH GAMEPLAY

FAR CRY 4 4K ULTRA SETTINGS	54fps AVERAGE FPS	43fps MINIMUM FPS



AMD Radeon™ R9 Nano Graphics Card

FORM FACTOR INNOVATION

175W SINGLE 8-PIN PCIe® CONNECTOR

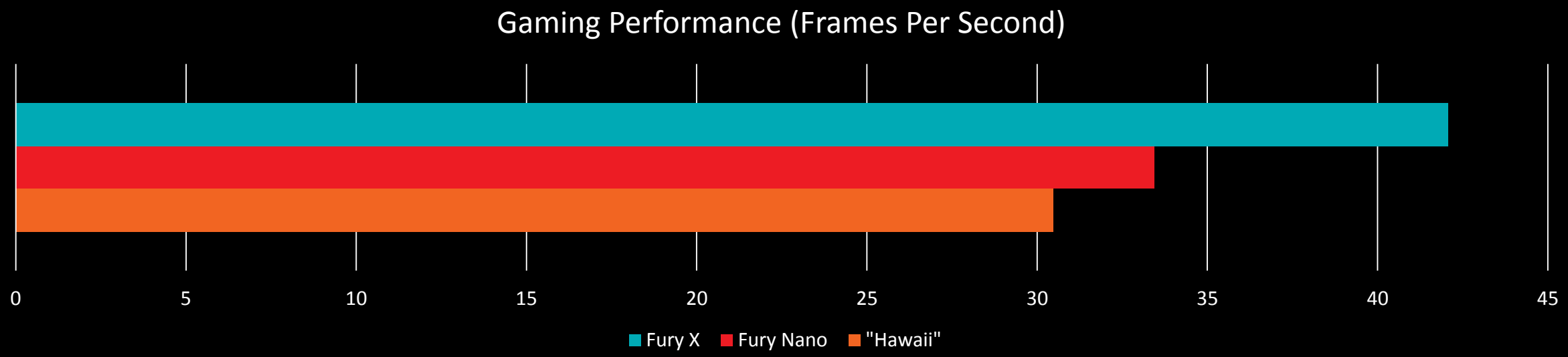
UP TO **2X*** PERFORMANCE DENSITY

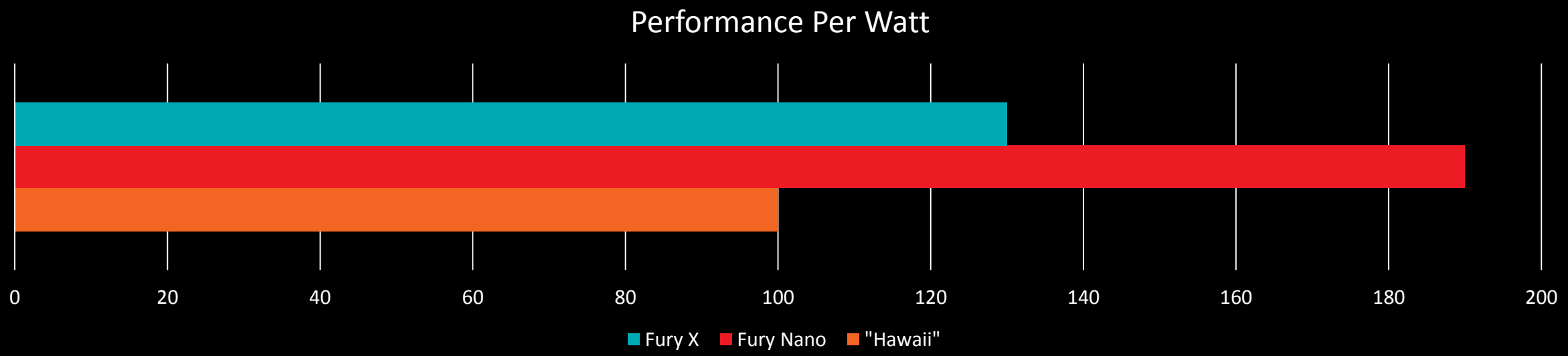
UP TO **2X*** PERFORMANCE PER WATT

*Compared to Radeon™ R9 290X graphics card
Board shot shown for illustration purposes only. Final board design may differ.

PERFOMANCE RESULTS[3]

AMD





Project Quantum

AMD

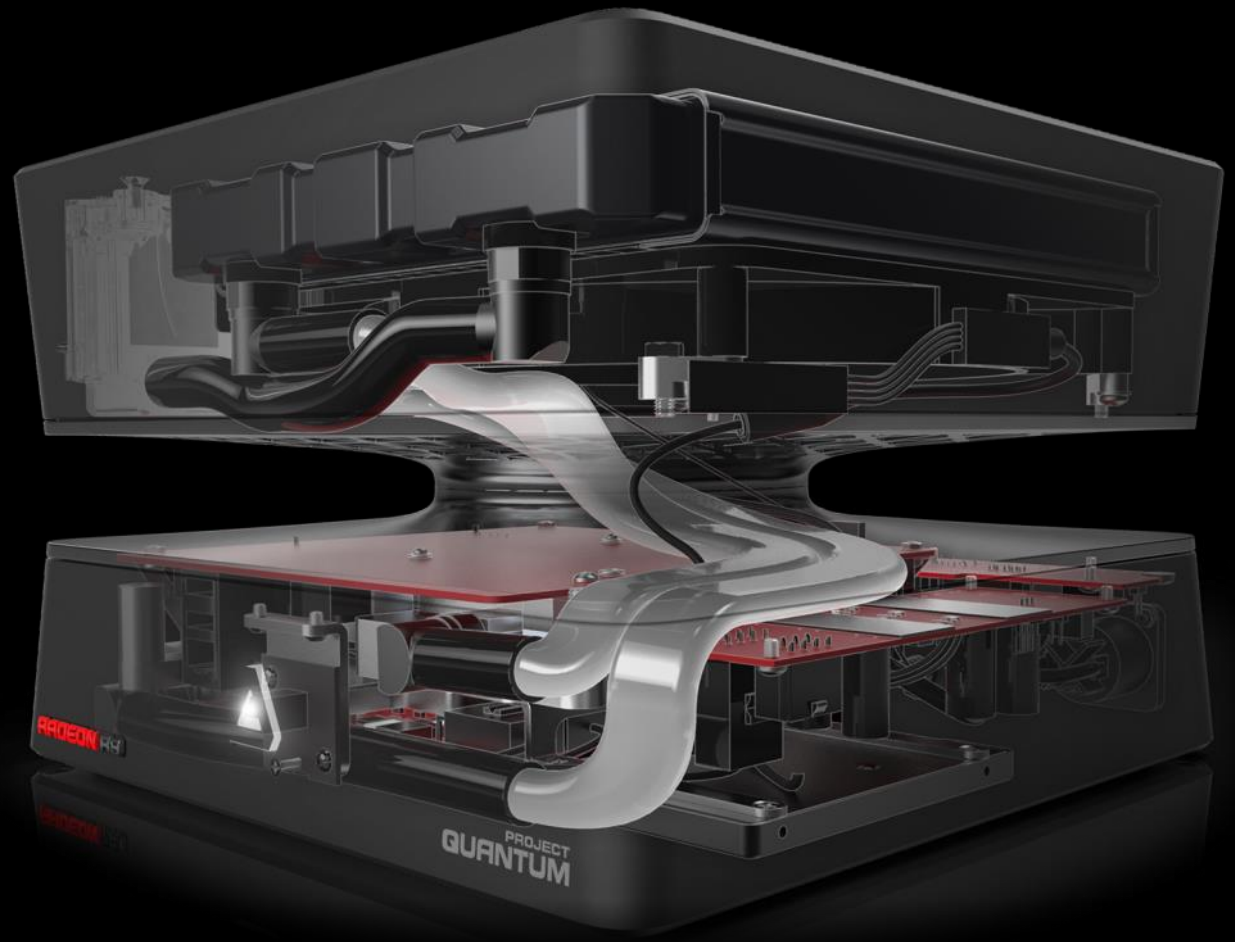
A NEW PC FORM FACTOR

- A unique approach to combine powerful hardware within a small form factor without compromises to thermals or acoustics
- Designed to deliver the best possible VR experiences with AMD LiquidVR™ technology

AMD

A NEW PC FORM FACTOR

- Powered by up to two Radeon™ R9 “Fiji” GPUs
- Fully liquid cooled system

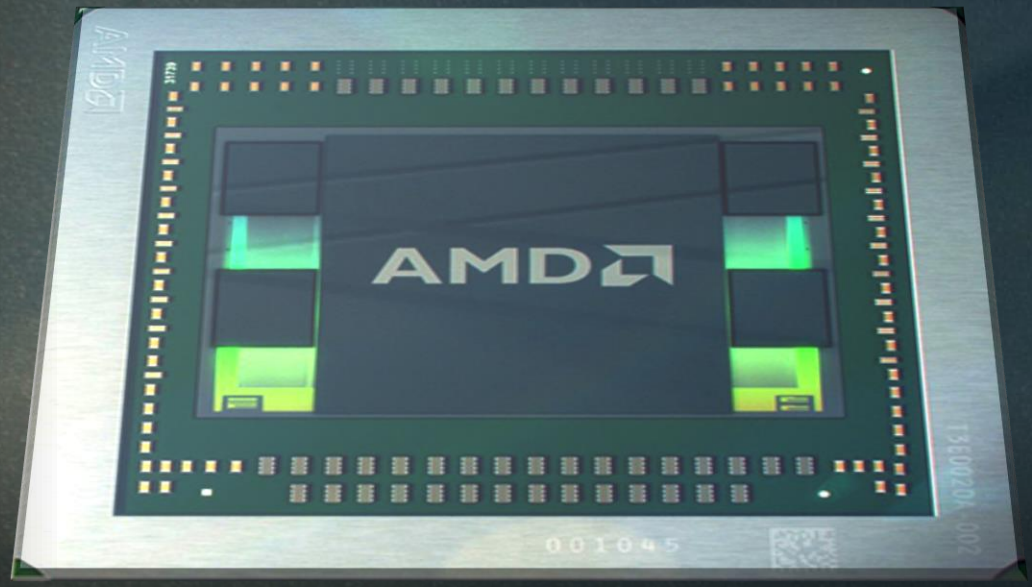


FURY

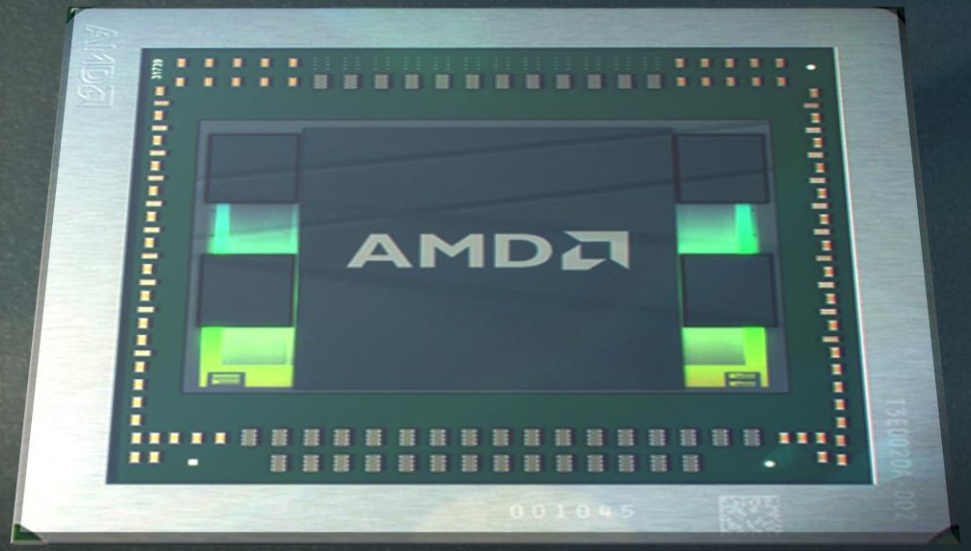
First with HBM

Enables Innovative Form Factors

Up to 2x Performance/Watt increase



Thank You



FOOTNOTES

1. Discrete AMD Radeon™ GPUs and AMD FirePro™ GPUs based on the Graphics Core Next architecture consist of multiple discrete execution engines known as a Compute Unit (“CU”). Each CU contains 64 shaders (“Stream Processors”) working in unison. GRT-5

2. Testing conducted by AMD engineering on the AMD Radeon™ R9 290X GPU vs. an HBM-based device. Data obtained through isolated direct measurement of GDDR5 and HBM power delivery rails at full memory utilization. Power efficiency calculated as GB/s of bandwidth delivered per watt of power consumed. AMD Radeon™ R9 290X (10.66 GB/s bandwidth per watt) and HBM-based device (35+ GB/s bandwidth per watt), AMD FX-8350, Gigabyte GA-990FX-UD5, 8GB DDR3-1866, Windows 8.1 x64 Professional, AMD Catalyst™ 15.20 Beta. HBM-1

3. Testing conducted by AMD engineering on the AMD Radeon™ R9 290X GPU vs. the AMD Radeon™ R9 Fury X GPU. Measured performance and power on Far Cry 4. System Configuration: Core i7-5960X (3001 MHz), AMD Catalyst™ 15.20 Beta.

DISCLAIMER & ATTRIBUTION

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