The Road to the AMD

“Fiji” GPU

Featuring Die Stacking and HBM Technology
“Fiji” Chip

DETAILED LOOK

- First high-volume interposer
- First 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
- 4096 Stream Processors
- 596 sq. mm. Engine

- 4GB High-Bandwidth Memory
- 4096-bit wide interface
- 512 GB/s Memory Bandwidth
DIE STACKING TECHNOLOGY

- Die stacking facilitates the integration of discrete dies
- 8.5 years of development by AMD and its technology partners
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
COMPLEXITY OF DIE STACKING

- **Business**: New technology partners, new business models, new roles for existing technology partners, new capital investments, no history for forecasting
- **Product**: Design utilization, CAD infrastructure, Define requirements, New IP
- **Test**: Wafer probe, KGD, Manufacturing thermal heads, Impact of gimbaling, Repair & Redundancy, DFT
- **Assembly manufacturing**: Big die challenges, Thin wafer handling, Die to die bonding thermal compression or reflow, Assembly order, New μbump technologies, EM, ESD, Underfill, Warpage control
- **Wafer Manufacturing**: TSVs, Carrier attach/detach, Thinning, μbump, Warpage control, DFM
- **Platform**: System characterization changes with integration, Definition of corners
- **Packaging**: Large and new materials to reduce warpage, BGA co-planarity
- **Thermals**: System solutions, Component solutions, Materials selection, Die thickness variation, Die co-planarity
- **Reliability**: New BOM, How is reliability proven with multiple vendor components with different requirements?
- **Quality**: How is quality distributed among technology partners? Who does burn in?
- **Yield & Cost**: DFY, Technology maturity time?

No single challenge is an issue but there are many, they are diverse, and they interact. Only functional prototypes can exercise these challenges.
8 YEARS OF DEVELOPMENT AND 6 MONTHS OF PRODUCT RAMP

- Three phases of the development over 8 years
  - 12 hours from 1\textsuperscript{st} sample to Windows™ booted and playing 3D games
  - Bring up complete in 5 days instead of 3 months
  - 6 months from ASIC wafer out to volume production

- Required collaboration with almost 20 different companies and government research organizations

- Four delivered final product: Amkor Technologies, ASE Group, SKhynix, and UMC

First Time Out: 2007 (100’s of samples)
Primary Learning: 2011 (<5000 of samples)
Product Readiness: Jul’14 (>5000 samples)

2014
PROTOTYPING ACTIVITIES OVER TIME

First Time Out

CPU + Mechanical DDR3
- AMD’s first interposer
- Large μBumps and TSVs
- Exercised complete test infrastructure
- Initial component reliability
- Initiated DA capability

2007 (100’s of samples) 2011
PROTOTYPEING ACTIVITIES OVER TIME

First Time Out

dGPU + Functional custom 2 die stacks of GDDR3
- 2nd interposer partner
- 1st OSAT partner
- Small μBump and TSVs
- Started definition of test flow with DRAM
- Continue component reliability
- Exercised DA capability

2007 (100’s of samples) 2011
PROTOTYPING ACTIVITIES OVER TIME

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

First Time Out

- CPU + D3 Mech.
- dGPU + G3

Primary Learning

- Began focus on enthusiast dGPUs
  - Larger solution 345mm² ASIC w/ 500mm² interposer
  - 3rd, 4th, and 5th interposer partners
  - 2nd and 3rd OSAT partners
  - Define assembly and test flow POR
  - Narrow the BOM
  - Define cooling solutions

Component reliability: TC | uHAST | HTS

- 2007 (100’s of samples)
- 2011 (<5000 of samples)
- Jul’14
AMD ACHIEVED FEASIBILITY IN 2012

- Multiple OSAT and Foundry partners
- Established POR for:
  - Business models
  - Wafer sort solutions
  - KGD solutions
  - System and component thermal solutions
  - Interposer features, process, and shipping
  - Manufacturing assembly flow
  - ESD solutions
  - Packaging materials
- Passed component reliability (TC, HTS, HAST, Precon)
- Passed board level reliability (TC, Bend, Vibration)
- Quality documents complete and implemented
- Yield >95%

Fully functional prototyping
## Prototyping Activities Over Time

### First Time Out
- CPU + D3 Mech.
- dGPU + G3

### Primary Learning
- 345mm² ASIC
- 500mm² IP
- dGPU

### Timeline
- **2007**
  - (100's of samples)
- **2011**
  - (<5000 of samples)
- **Jul'14**
  - Cost Down
  - ESD | BLRT | Sort
  - PwrCyc | uBump EM | TSV EM/SM
  - Component reliability: TC | uHAST | HTS
PROTOTYPING ACTIVITIES OVER TIME

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

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

First Time Out

- 2007 (100's of samples)
- CPU + D3 Mech.
- dGPU + G3

Primary Learning

- 2011 (<5000 of samples)
- 345mm² ASIC
- 500mm² IP
- dGPU
- Routing for Daisy Chain Stack
- Interposer
- Stiffener

- 2014 Jul'14 (>5000 samples)
- 502mm² ASIC
- 818mm² IP
- dGPU

Product Readiness

- 2014
- 592mm² ASIC
- 1011mm² IP
- "Fiji" Replica
- Cost Down
- PwrCyc
- uBump EM
- TSV EM/SM
- Component reliability: TC
- uHAST
- HTS

(<5000 of samples)
HIGH-BANDWIDTH MEMORY

- Initiated with several DRAM partners 7 years ago
- SKhynix is in production supporting “Fiji”

Benefits
- 4096-bit memory interface with four stacks creating 512GB/s of bandwidth
- 60% higher memory bandwidth for 60% less power than GDDR5
- 4X Bandwidth per watt improvement from Radeon™ R9 290X

Also required functional prototyping
HIGH-BANDWIDTH MEMORY

- AMD and SKhynix developed a vehicle to exercise the uBump interface and mission mode test HBM
- Controller dies implement AMD MC and DFT
- Mission mode testing of uBump interface
EFFICIENT DESIGN

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

PCB area occupied by ASIC with HBM

1.6X shorter

2X shorter

~3X reduction in PCB Footprint with HBM
AMD RADEON™ R9 FURY SERIES GRAPHICS
ENGINEERED FOR 4K AND VR GAMING

AMD RADEON™ R9 FURY X
AMD RADEON™ R9 NANO
AMD RADEON™ R9 FURY
EVOLUTION OF THE 4K GAMING PC

AMD RADEON™ R9 290X

AMD RADEON™ R9 FURY X

AMD RADEON™ R9 NANO
RADEON™ R9 NANO

40% SHORTER THAN RADEON™ R9 290X GRAPHICS

6” PRINTED CIRCUIT BOARD
TARGET OPERATING TEMPERATURE

75°C

FASTER, COOLER THAN RADEON™
R9 290X GRAPHICS™
RADEON™ R9 NANO

LIBRARY QUIET

16 dBa QUIETER THAN RADEON™ R9 290X GRAPHICS™

*See footnote 3
RADEON™ R9 Nano

Up to **2X** PERFORMANCE PER WATT

Up to **2X** PERFORMANCE DENSITY™

**175W** TYPICAL BOARD POWER

*Compared to the Radeon™ R9 290X graphics card. See footnote 4
**Compared to the Radeon™ R9 290X graphics card. See footnote 5
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.

ATTRIBUTIONS

© 2015 Advanced Micro Devices, Inc. All rights reserved. AMD, the AMD Arrow logo, Radeon, and combinations thereof are trademarks of Advanced Micro Devices, Inc.

DirectX and Microsoft are registered trademarks of Microsoft Corporation in the US and other countries.

Other names are for informational purposes only and may be trademarks of their respective owners.
FOOTNOTES

Required for all AMD Radeon™ graphics messaging: Additional hardware (e.g. Blu-ray drive, 4K monitor, TV tuner, wirelessly enabled HDTV) and/or software (e.g. multimedia applications) are required for the full enablement of some features. HD Video display requires an HD video source. Not all features may be supported on all components or systems - check with your component or system manufacturer for specific model capabilities and supported technologies.

1. 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 (42.66 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

2. Based on the product design, the Radeon™ R9 Nano is defined with an operating temperature target of 75°C while the Radeon™ R9 290X is defined with an operating temperature target of 95°C.

3. Based on the product design, the Radeon™ R9 Nano is defined with a fan acoustic target of 42dBA while the Radeon™ R9 290X is defined with a fan acoustic target of 58dBA.

4. Testing conducted by AMD Engineering on optimized AMD reference systems. PC manufacturers may vary configurations yielding different results. Far Cry 4 at 3840x2180, Ultra High preset, SMAA, 0XAF is used to simulate GPU performance; the Radeon™ R9 Nano on the system using the Intel® Core™ i7-5960X 3.0GHz processor, 16GB (4x4GB) DDR4 2666 MHz memory, Windows 10 64-bit, and AMD Catalyst Driver 15.201 scored 0.2169 fps/watt while the Radeon™ R9 290X on the same system and AMD Catalyst Driver 15.20 scored 0.1088 fps/watt.

5. Testing conducted by AMD Engineering on optimized AMD reference systems. PC manufacturers may vary configurations yielding different results. Far Cry 4 at 3840x2180, Ultra High preset, SMAA, 0XAF is used to simulate GPU performance; the Radeon™ R9 Nano on the system using the Intel® Core™ i7-5960X 3.0GHz processor, 16GB (4x4GB) DDR4 2666 MHz memory, Windows 10 64-bit, and AMD Catalyst Driver 15.201 scored 0.2498 fps/mm while the Radeon™ R9 290X on the same system and AMD Catalyst Driver 15.20 scored 0.0989 fps/mm.

6. Based on the memory bandwidth of the AMD Radeon™ R9 290X with a 1250MHz 512-bit GDDR5 interface (320GB/s) vs. AMD Radeon™ R9 Fury and R9 Fury X featuring HBM with a 500MHz 4096-bit interface (512GB/s). HBM-4

7. Testing conducted by AMD engineering on the AMD Radeon™ R9 290X GPU vs. the AMD Radeon™ R9 Fury X GPU. Data obtained through isolated direct measurement of GDDR5 and HBM power delivery rails at full memory utilization. AMD Radeon™ R9 290X and R9 Fury X GPU, AMD FX-8350, Gigabyte GA-990FX-UD5, 8GB DDR3-1866, Windows 8.1 x64 Professional, AMD Catalyst™ 15.20 Beta. HBM-3

8. 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