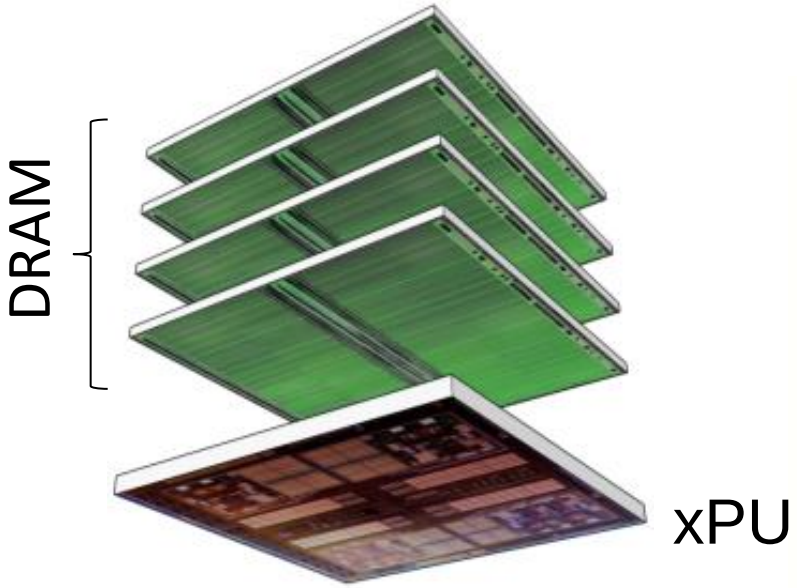
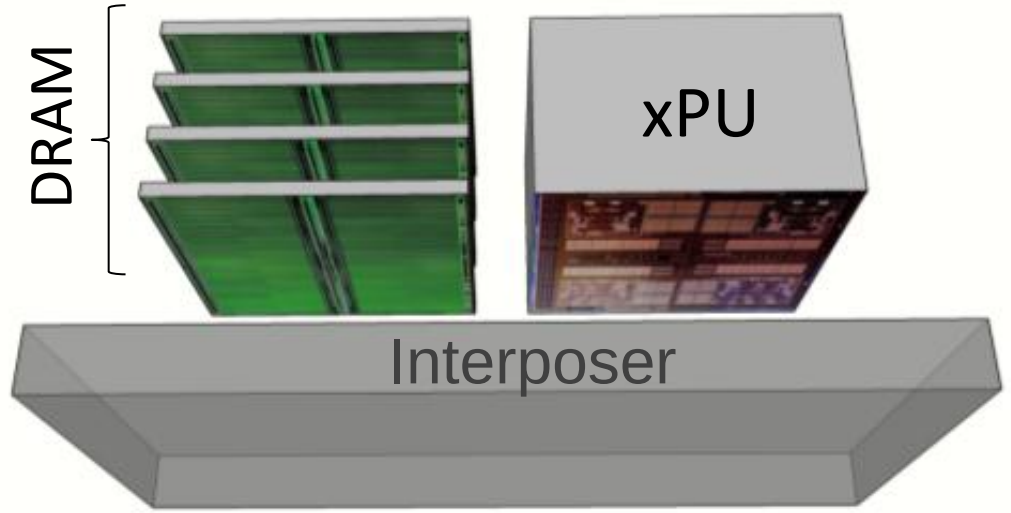
AMD

EMERGING TECH PANEL: 3D MEMORY INTEGRATION

Gabriel Loh
AMD Research

WHAT IS DIE-STACKING?



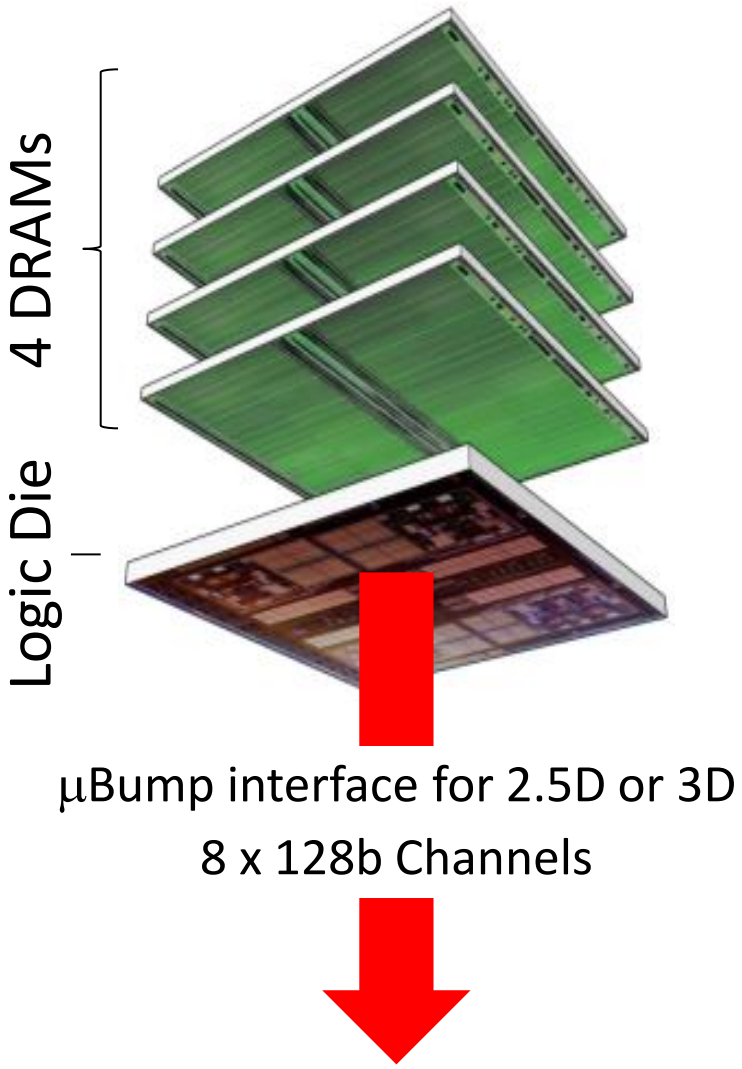


VERTICAL STACKING (3D) **INTERPOSER STACKING (2.5D)**

- Mix and match different technologies (e.g., CMOS, DRAM)
- Replace costly off-package interconnects with short, fast, low-energy in-package metal routes
- Increase integration density

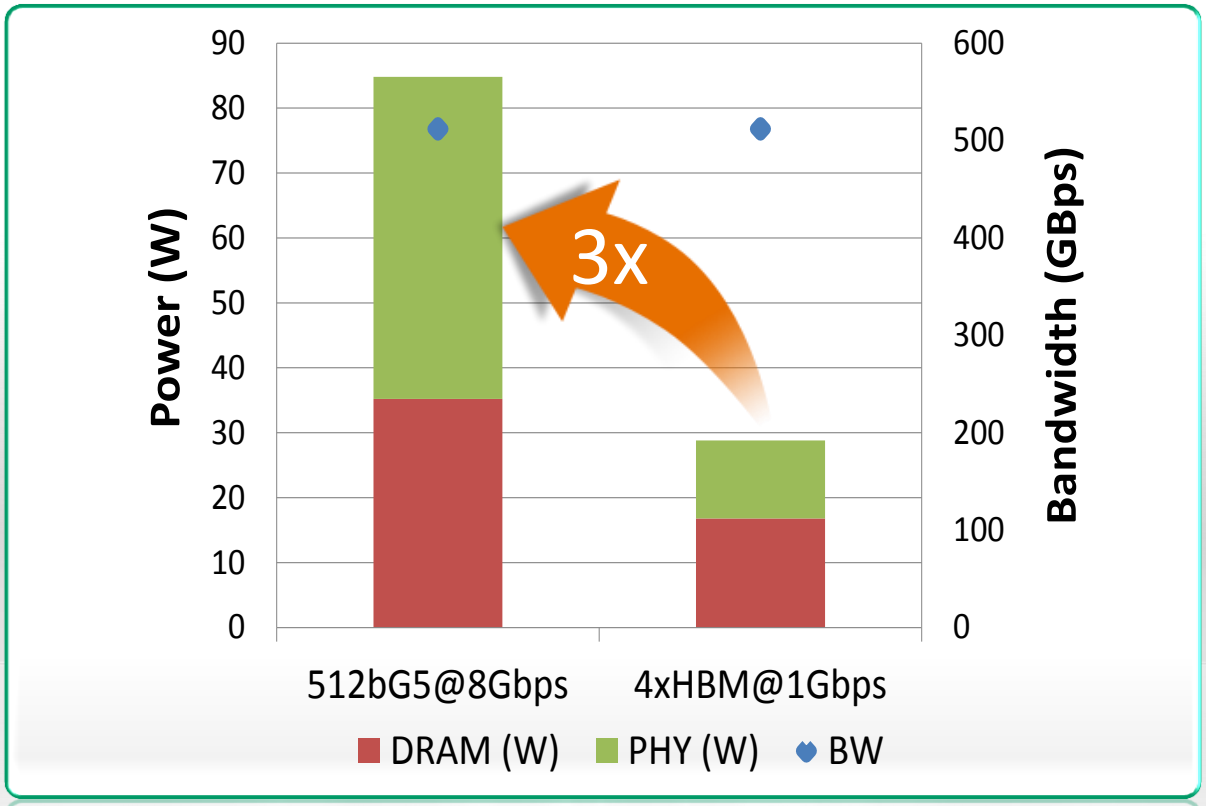
1ST GENERATION HBM IS SAMPLING!

HTTP://WWW.JEDEC.ORG/STANDARDS-DOCUMENTS/DOCS/JESD235



Voltage (I/O, VPP)	1.2V; 2.5V
Data Rate per pin	800 – 1.2 MT/s
DRAM Die Density	2Gb
DRAM Dies per Stack	4
Interface Width	1024
DQs/Channel	128
Channels/Die; Channels/Stack	2; 8
Address/Command Interface	DDR
Data Bus Clocking	Source synchronous; uni-directional/4byte
On Die DLL	No
Data Bus Termination – On die	No
Address/Cmd Termination – On die	No
DQ Training Mode	N/A
Page Size	2KB
Banks/Channel	8
DRAM Prefetch; Burst Length	256 bits; 2

IMPROVING BW/WATT WITH DIE-STACKED MEMORY



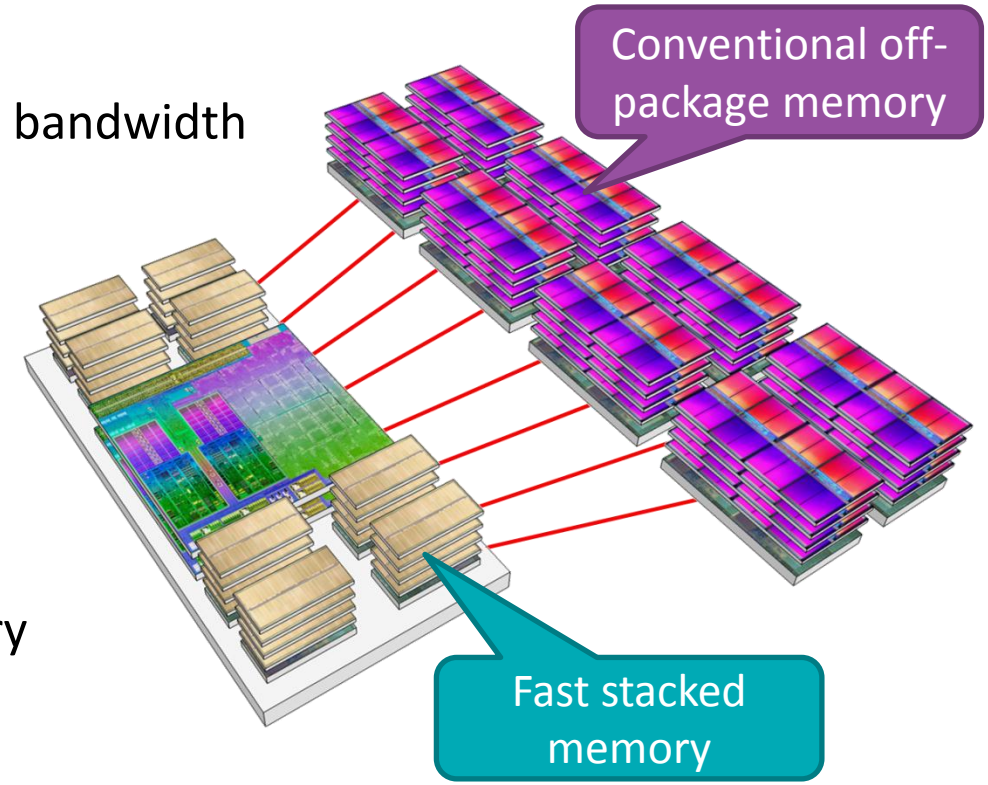
Source: AMD

- Die stacking improves the proximity of the DRAM to Compute
- Dense and fine pitch interconnect enables simple low power interfaces as well as fine grain power control of the DRAM

WHAT THIS MEANS FOR HPC

AMD

- In-package memory to deliver high bandwidth at low power cost
- But… in-package memory does not provide enough capacity
  - Still need conventional memory for capacity
- Results in a heterogeneous memory organization/hierarchy



- How does one make small-fast + large-slow memory act like a single, uniform large+fast memory?
  - Hardware? Software? Combination?

PANEL QUESTIONS

AMD

- Q1: What is most needed to bring to market?
  - Hardware: technology is there, now need more volume to drive costs/risks down
  - Software: may be more evolutionary, multiple usage models
- Q2: What markets will use it first? Will it be mainstream?
  - Where bandwidth needs are the highest (GPU)
- Q3: What supporting technologies are required?
  - For wide adoption, software support needs to be there
  - System software development dependent on hardware availability
  - Applications optimization dependent on system software, runtimes, libraries, etc.
- Q4: What partners would help bring to market sooner?
  - SoC vendors using more stacked memory → volume
    - Enabled/enhanced by common standards (e.g., JEDEC HBM)
    - More platforms → more software development

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