# 3D Packaging for Memory Application

## (NCCAVS Joint User Group Meeting)

Ming Li, Ph.D.

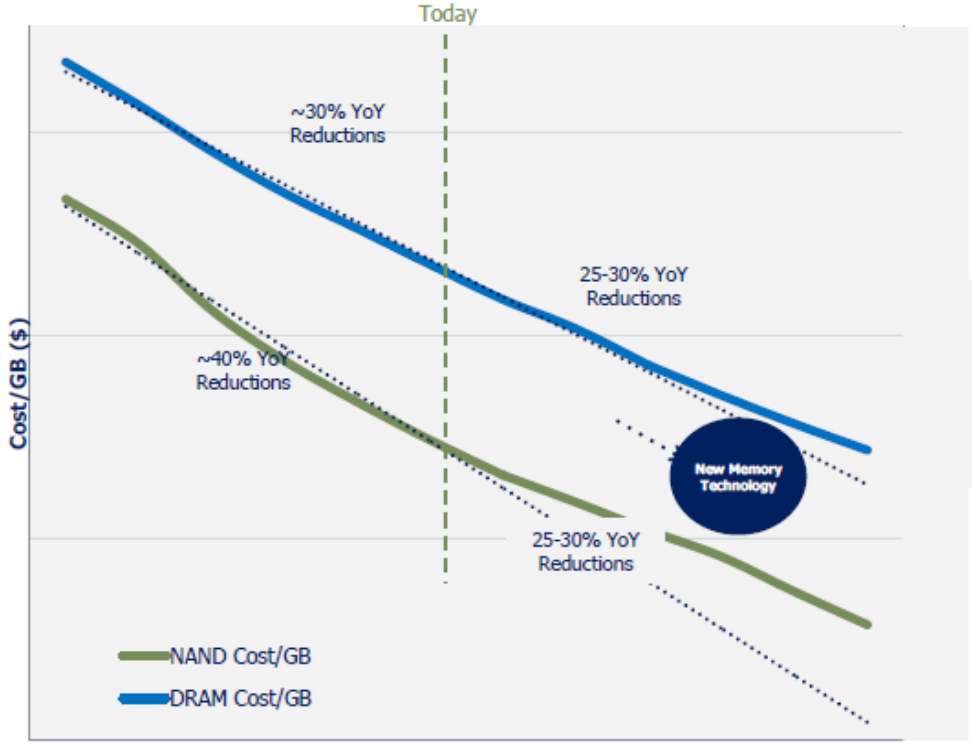
June 11, 2013

Rambus

# Outline

- Motivations for 3D Packaging Technology Adoption in Memory Space
  - Memory Cost Scaling Challenge
  - Power Efficiency Requirement
  - Memory Sub-system Performance Demand
- Traditional 3D Packaging Technologies for Memory Application
  - System-in-Package (SiP)
  - Package-on-Package (PoP)
- Emerging TSV Based 3D Packaging Technologies for Memory Application
  - Mobile Wide IO DRAM
  - Hybrid Memory Cube (HMC)
  - High Bandwidth Memory (HBM)
- Rambus 3D Package for Logic/Memory Application
- Summary

# Memory Cost Scaling Challenge

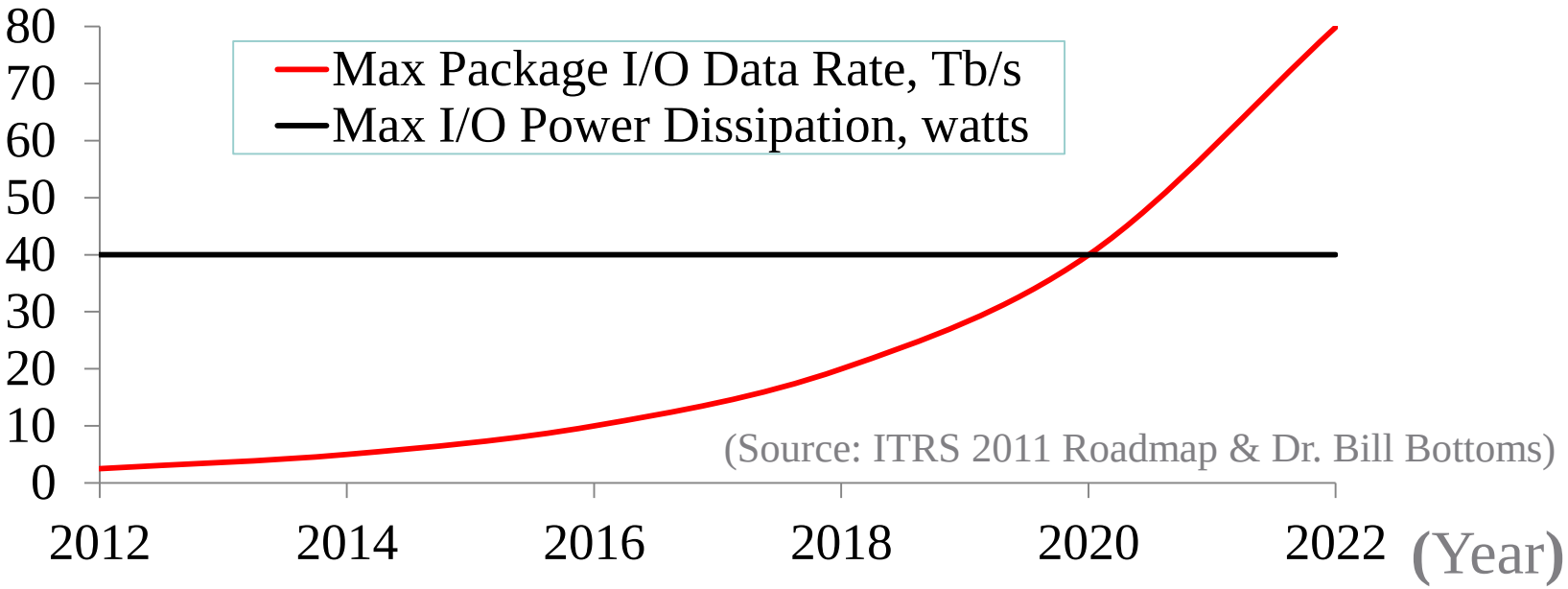



**Memory Cost Scaling Over Time**

- **Physical and Electrical Materials Limitations**
- **Equipment Capability**
- **Cost and Complexity**

(Source: Mike Black, Micron, EDPS 2013)

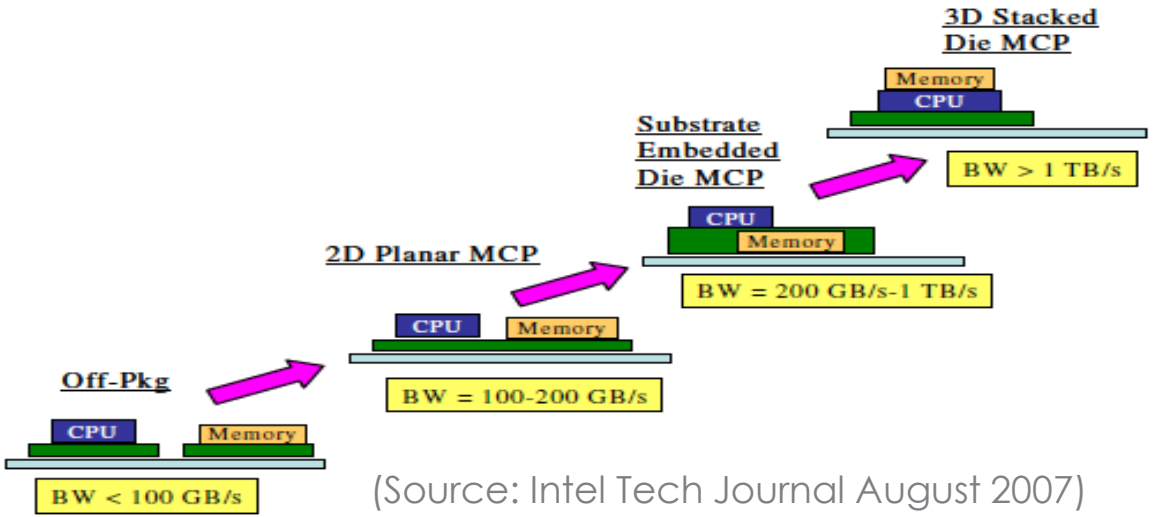
# Power Efficiency Requirement




- Scaling of I/O date rate with constant I/O power dissipation
- Even more challenging for computing memory subsystem memory bandwidth (BW) scaling

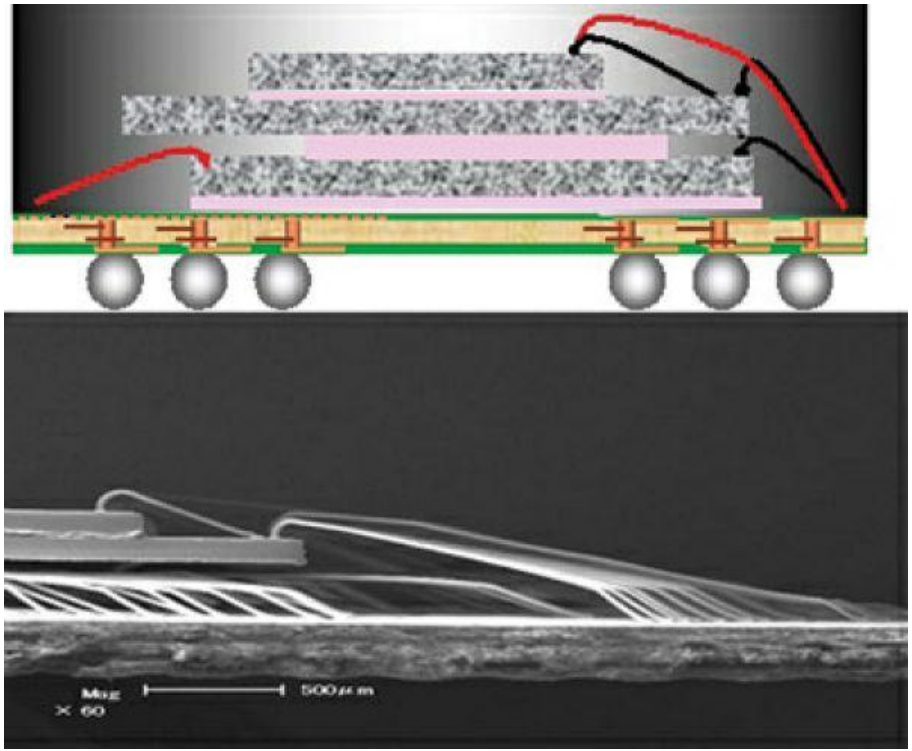
# Memory BW Demand

- Packaging and interconnect technology vital in defining memory sub-system performance
- Traditional off-package interconnection between CPU and memory chip not going to scale, trends: on-package interconnection and 2.5D/3D interconnection




(Source: Intel Tech Journal August 2007)

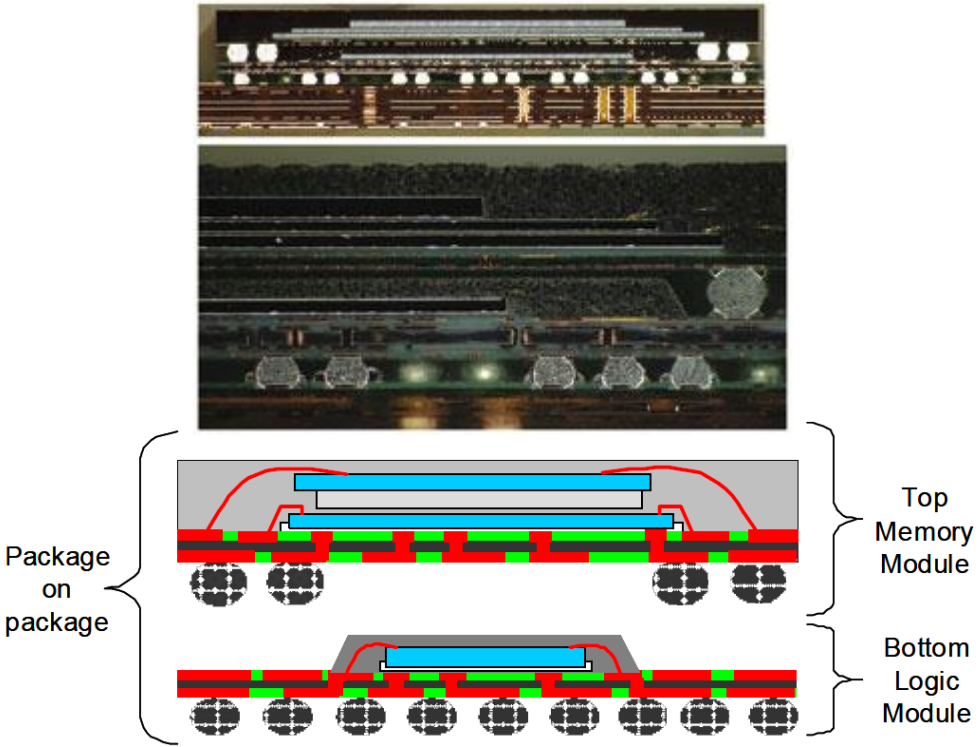
# Traditional 3D Packaging for Memory Application - SiP




(Source: Internet)

- System-in-Package (SiP)
  - Die stacking using wire bond (WB) interconnects
  - Same memory die stacking for memory capacity (NAND or DRAM)
  - Logic/memory die stacking for memory bandwidth (BW)
  - Low cost solution
  - Limited BW scalability due to limited WB interconnect density
  - Know-Good-Die (KGD) concern

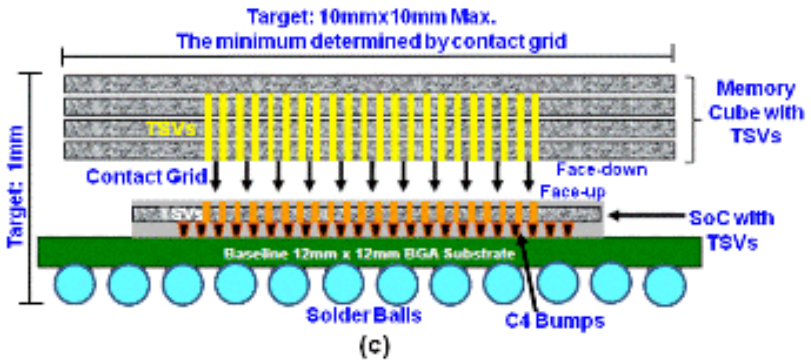
# Traditional 3D Packaging for Memory Application - PoP



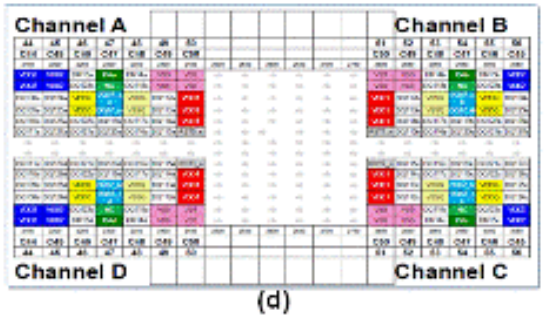

(Source: Internet)

- Package-on-Package (PoP)
  - Top memory package
  - Bottom logic package
  - Logic and memory packages separated for logistic control and KGD
  - Standard memory interface between top and bottom
  - Limited BW scalability due to limited interface memory bus width
  - Mainly for mobile application

# TSV Based 3D Packaging Technology for Memory Application – Wide IO DRAM



(c)



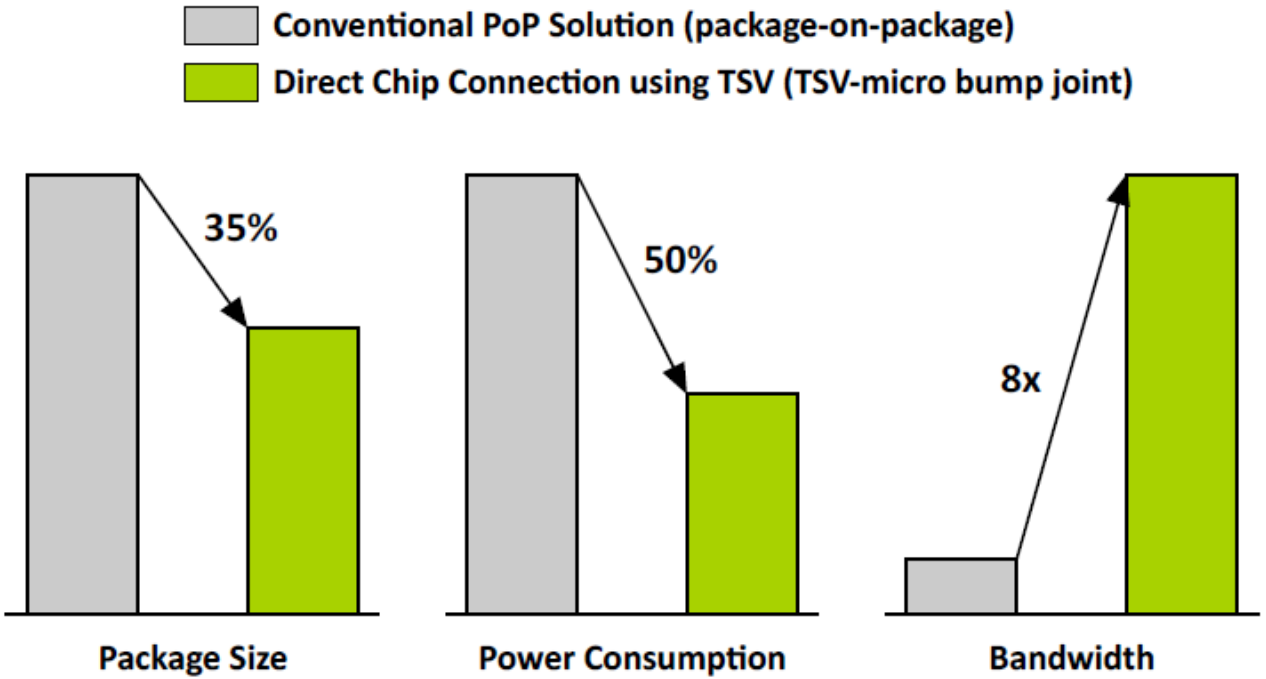
(d)

| | WideIO | LPDDR3 |
|---|---|---|
| Power | 1.2 volts | 1.2 volts |
| Speed | 200 Mbps | 1600 Mbps |
| Data width | 512 bits | 32 or 64 bits |
| Data rate | single | double |
| Maximum channel bandwidth | 12.8 Gbps | @32 bits -- 6.4 Gbps |

- Low power for mobile application
- Small formfactor
  - Thin profile
- High cost
- Complicate business model
  - Yield loss
  - Scrap management

(Source: Internet)

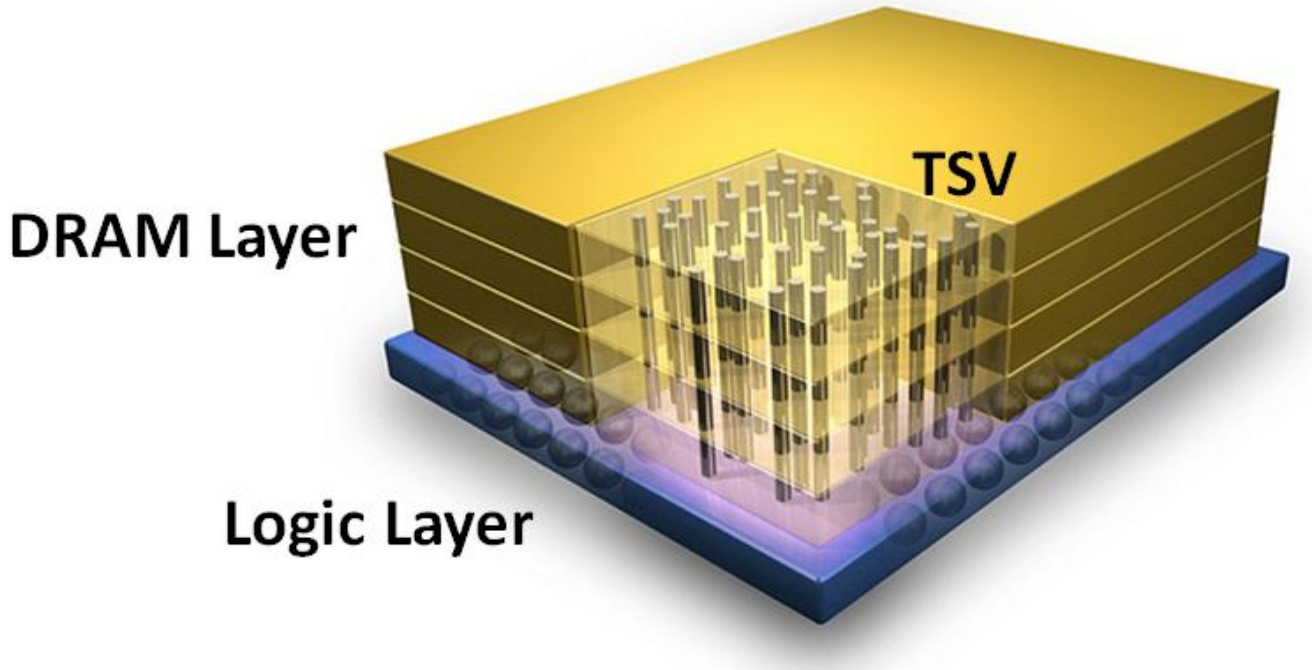
# Wide IO DRAM Stacking vs. PoP




Note: Wide IO DRAM (512bits @ 200MHz SDR)

(Source: Samsung)

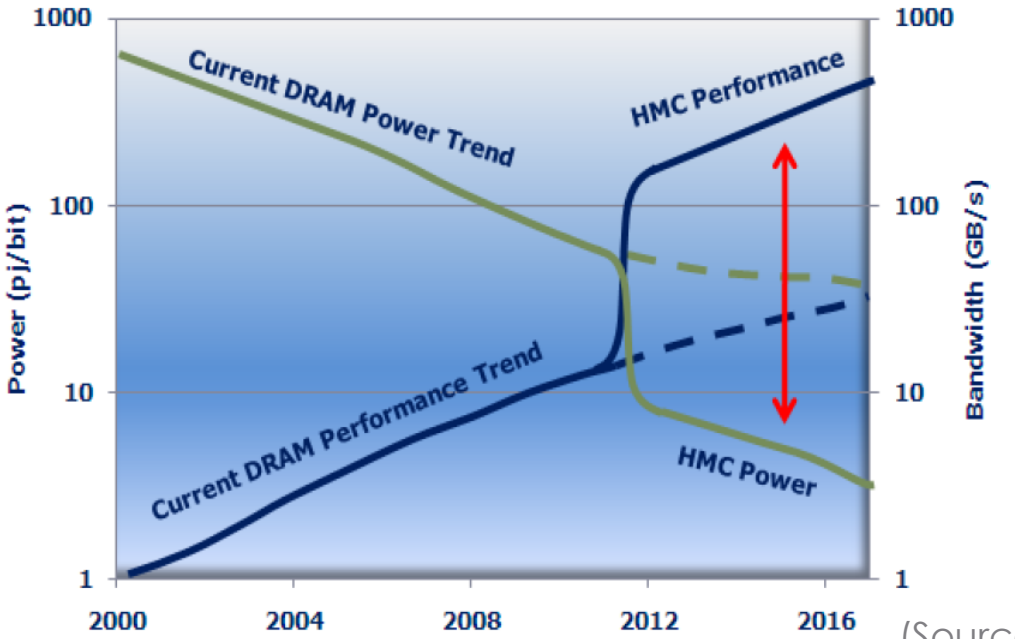
# TSV Based 3D Packaging Technology for Memory Application – Hybrid Memory Cube (HMC)




(Source: HMCC, Micron)

# Hybrid Memory Cube (HMC)

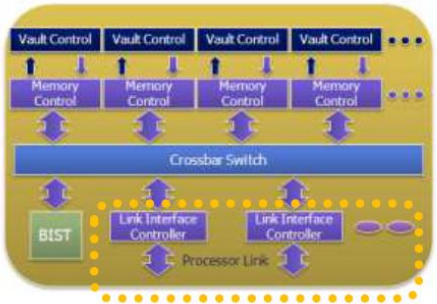
Fast process logic and advanced DRAM design in one optimized package



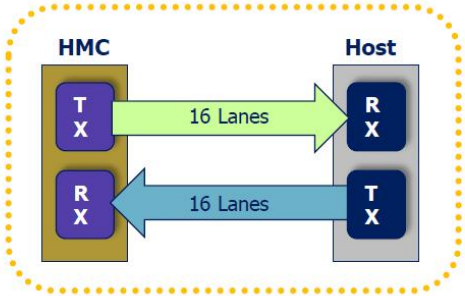

- Power Efficient
- Smaller Footprint
- Increased Bandwidth
- Reduced Latency

(Source: MemCon Memory Conference 2012)

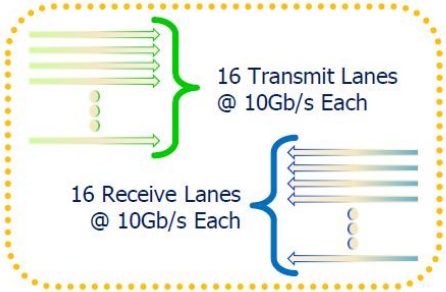
# HMC Link Controller Interface




HMC (Gen2) has Four Links




Each Link has 32 differential Lanes




(*) HMC-SR supports up to 15Gbps SERDES options

Designed with Off-the-Shelf, High Speed SerDes Interface

**Link Interface Examples**

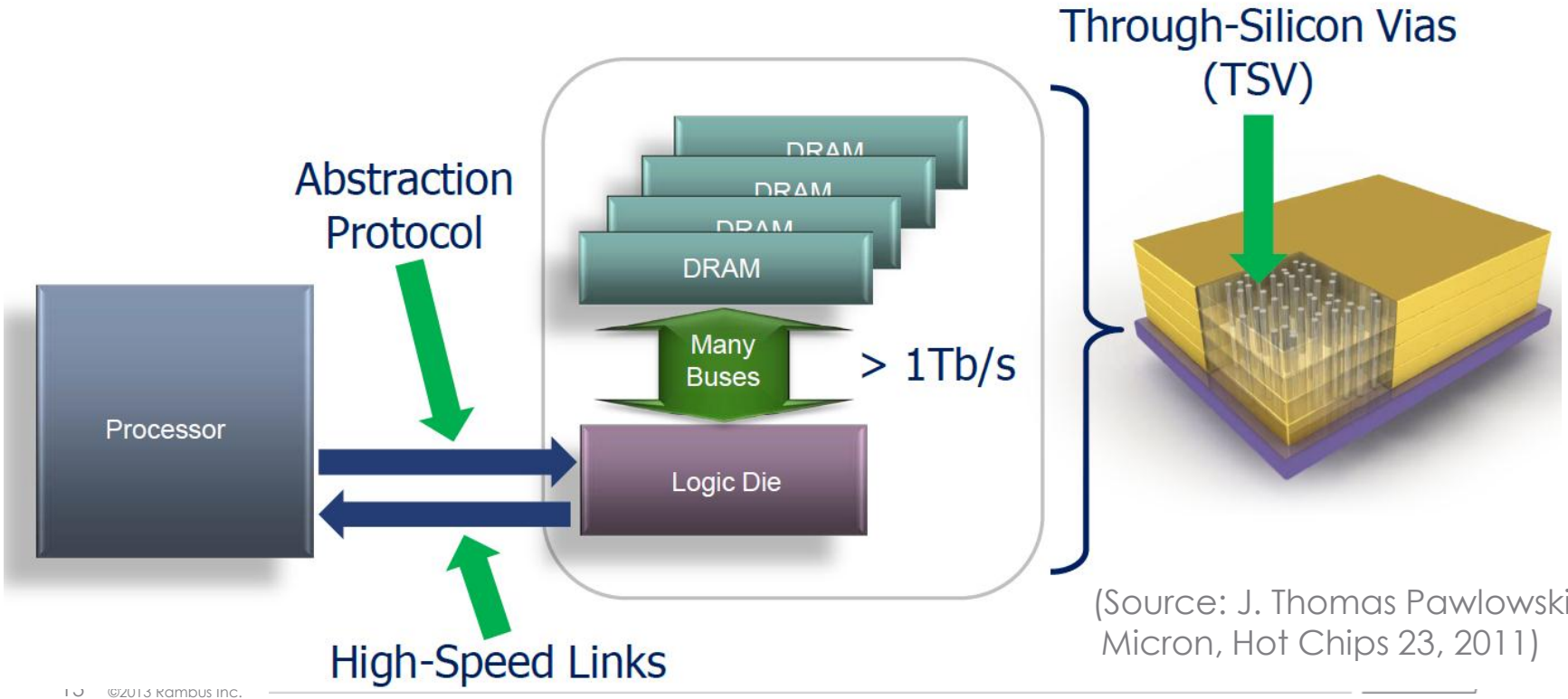
**4 Link Example (160GB/s)**
- 10Gb/s per lane
- 32 lanes per link (320Gb/s = 40GB/s)
  - 16 TX and 16 RX
- 4 Links (40GB/s x 4) = **160GB/s**

(*) Aggregate DRAM peak bandwidth remains 160GB/s

- Up to 15X the bandwidth of a DDR3 module
- 70% less energy usage per bit than existing technologies
- Occupying nearly 90% less space than today's RDIMMs

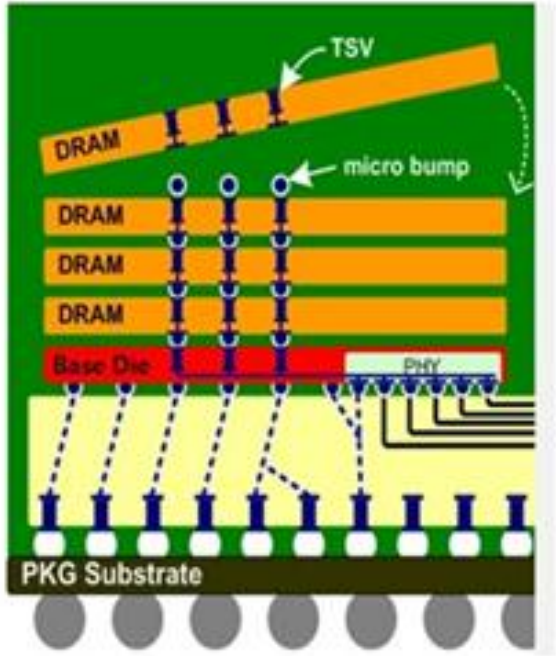
(Source: Mike Black, Micron, EDPS 2013)

# HMC Application




(Source: J. Thomas Pawlowski, Micron, Hot Chips 23, 2011)

# TSV Based 3D Packaging Technology for Memory Application - High Bandwidth Memory (HBM)




| ITEM | TARGET |
|---|---|
| Burst Length | 2, 4 |
| Stack Density | 1GByte per stack (2Gbit per slice) |
| Channel / Slice | 2 |
| Banks / Channel | 8 |
| IO / Channel | 128 |
| Prefetch / Channel | 32B (128x2bit) |
| Channels / Stack | 8 |
| Total TSV Data IO Width | 1024 |
| Clock Speed | 500MHz |
| Peak Read BW / Stack | 128 GB/s |
| Page Size | 2KB |
| Data Parity | 1 bit / 32 bit |
| DRAM Core Voltage | 1.2V |
| Logic Buffer IO Voltage | 1.2V |

(Source: SK/Hynix, Jan 2013)

# High Bandwidth Memory (HBM)

## Major Features of HBM

### Wide I/O Stacked DRAM with TSV

| GDDR5 | Archi-tecture | HBM |
|---|---|---|
|   | |   |
| x32 | IO | x 1024 |
| 7Gbps | Data Rate | 1Gbps |
| 1.5V | Voltage | 1.2V |

(Source: SK/Hynix, Jan 2013)

# Markets for TSV Based 3D Packaging Technologies

| Market | SmartPhone | Tablet | Networking | Graphics |
|---|---|---|---|---|
| Processor | Apps Processor | Apps Processor | Networking Proc | Graphic Proc |
| Power | 1-2W | 1-5W | 20W + | 20W + |
| Memory Type | Wide I/Ox | Wide I/Ox or LPDDRx | HMC / HBM | HBM |
| Memory Size | 2 → 4 GB | 4 → 8 → 16 GB | 4 → 8 → 16 GB | 4 → 16 GB |
| Interface | Wide I/O2 | Wide I/O or DDR | SerDes / Parallel | Parallel |
| I/O | 1000 | 1000 or 500 | <500 / >1000 | 1600 + |
| Min Bump Pitch | 40x40µm rows | 40x50µm rows or 80µm rows | 1mm / 96x55µm array | 96x55µm array, staggered rows |
| Packaging | 3D | 2.5D medium L/S density or 3D with heat management | Off chip memory or 2.5D high density interposers | 2.5D high density interposers |




(Source: Internet)

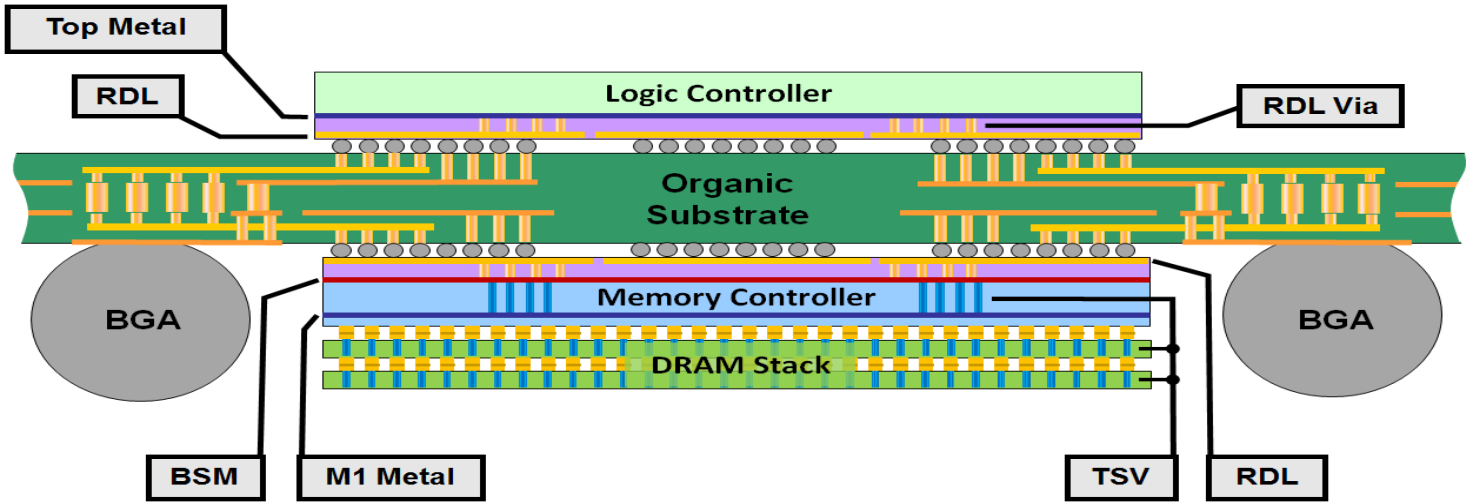
# Rambus 3D Package for Logic/Memory - Introduction

- CPU/GPU performance often constrained by both insufficient memory bandwidth and excessive memory power
- Stacked CPU/GPU and DRAM (using TSV) poses significant challenges to thermal management and power delivery
- One solution is double-sided flip-chip package with a processor on top and stacked DRAMs on bottom
- Middle-ground approach between traditional package-to-package, on-PCB solutions and pure-3D integration
- Processor interfaces with memory directly through package
- Very high IO density and low-loss interconnect paths
- Thermally decoupled processor and memory devices

(Package co-design & optimization: Secker D.; et al., ECTC 2012)

# 3D Package Physical Architecture

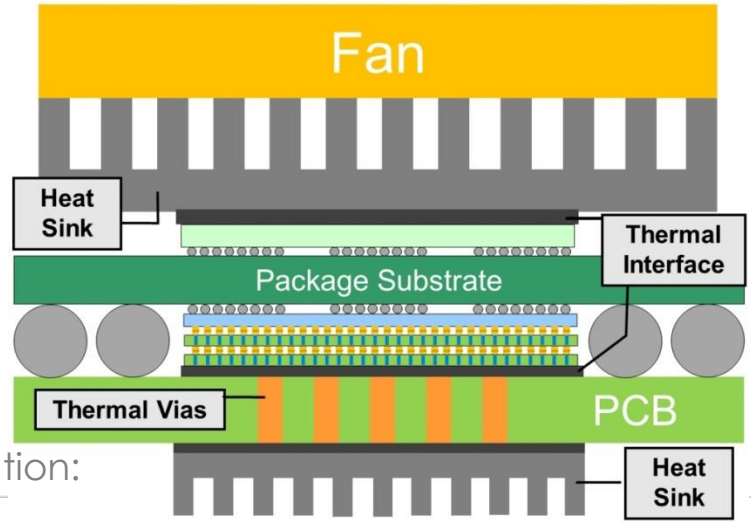
- Organic Substrate and Assortment of BEOL Technologies
  - Controller and DRAM stack share a 5-2-5 organic substrate
  - RDL on Controller and DRAM
  - Disaggregated DRAM with fine-pitch TSV's
  - 256 GB/s peak DQ bandwidth across 800 diff. pairs (4 Gb/s per pair)




(Package co-design & optimization: Secker D.; et al., ECTC 2012)

# Thermal Design

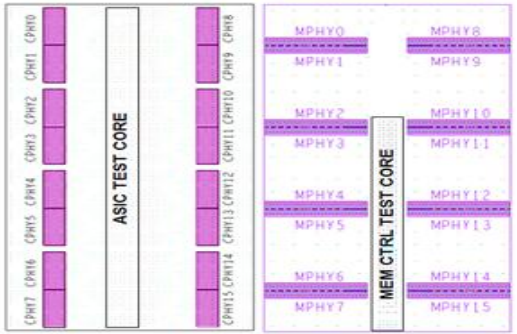
- High power of stacked memory with high-performance processor creates thermal management issue in 3D IC
- Thermal isolation achieved mounting components on opposite sides of package substrate
- Memory stack cooled using combination of thermal interface material, PCB thermal vias and back-side heat sink




(Package co-design & optimization: Secker D., et al., ECTC 2012)

# 3D Package Implementation

**Controller and DRAM Floorplans**

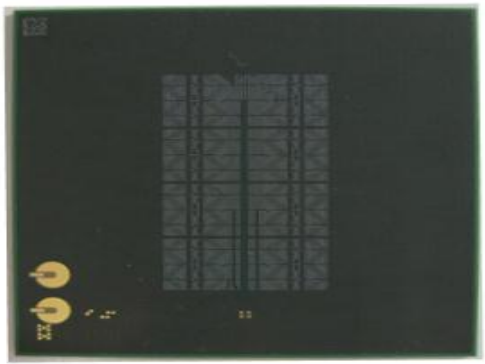



**Controller Die**

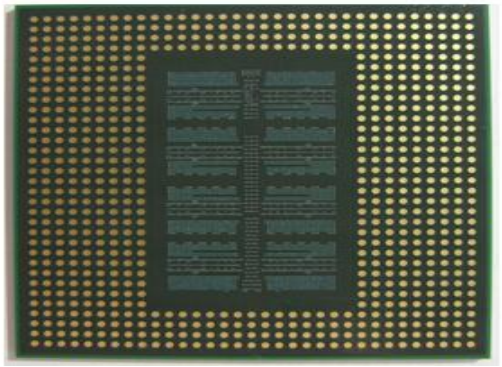
- 12.3 x 21.6mm
- 200um C4 pitch

**Organic Package Substrate**

- 35x35mm outline
- Thin-core 5-2-5 buildup
- 718 BGA (1mm pitch)
- ~ 8800 C4 (total)



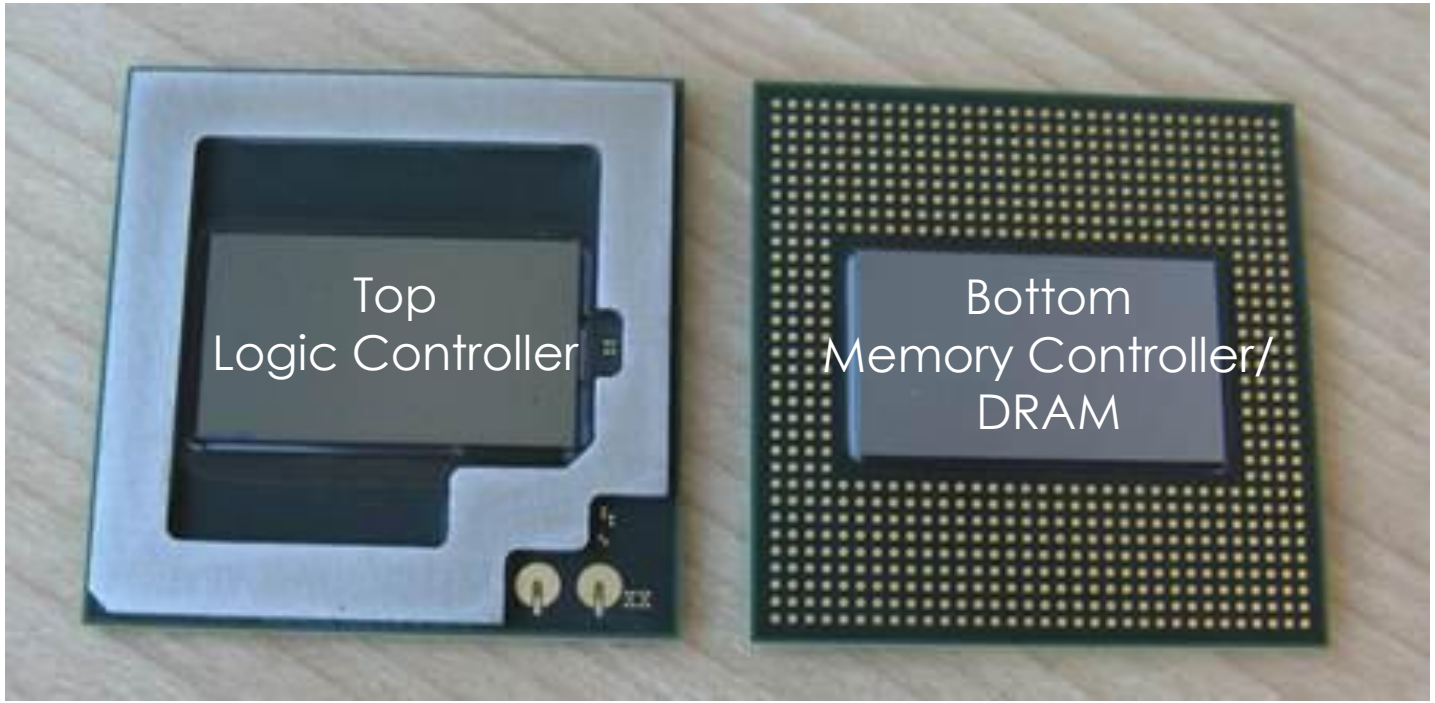
**Top View**



**Bottom View**

(A 256GB/s Memory Subsystem Built Using a Double-Sided IC Package with a Memory Controller and 3D-Stacked DRAM: Best S.; et al., DesignCon 2013)

# Rambus 3D Package for Logic/Memory



# Summary

- Cost scaling, power efficiency and bandwidth are three main factors that drive the adoption 3D packaging technology for memory application.
- Traditional 3D packaging technologies like SiP and PoP have reached the scaling limit for performance and power efficiency.
- Emerging TSV based 3D packaging technology shows great potential for scaling but many issues have to be resolved before it could be a mature technology.
- Rambus 3D packaging technology for logic/memory application bridges both mature technologies and advanced 3D IC technology and shows great potential for immediate implementation.

# Thank you

rambus.com

Rambus