3D XP: What the Hell?!!

Dave Eggleston
Making Sense of 3D XP

• **WHY** are you surprised?
• **WHAT** is it (really)?
• **XP Array Limitations: The Litho Problem**
• **HOW** will it be used?
• **3D XP strengths & weaknesses**
## Leading-Edge Technology Status

**DRAM**
- Continued strong progress for 20nm yield at Hiroshima and Inotera facilities
- 1Xnm development underway in Hiroshima and 1Y/1Znm in Boise

**NAND**
- 16nm TLC NAND now ramping in Singapore
- Micron 1st generation 3D NAND on track for production in Singapore mid 2015
- Second generation under development in Boise

**Package Technology**
- 3D NAND package technology continues to mature, currently manufacturing HMC generation 2
- R&D focus on HMC generation 3 enablement for even higher density and bandwidth

**New Memory Technology**
- Multiple paths under active development for storage class memory enablement
- Targeting 2015 and 2017 for manufacturing introductions of next new memory technologies
Innovation Roadmap

- Increased focus on DRAM technology position driving faster introduction cadence
- Enable volume 3D NAND manufacturing capability through 2015
- 3D NAND packaging technology enablement for multiple differentiated opportunities
- Establish disruptive new memory technology and position for ramp in 2016

Source: S. DeBoer; Micron Analyst Day, Feb 2015
New Memory Technologies

- Replacement technologies are only one aspect
- Gap between memory and storage is another
- Need increased freedom to balance important system parameters

New Memory A is:
- DRAM-like
- Low latency
- Relatively high endurance
- Highly volatile(!)
- Relatively high cost
Switching Mechanisms

**Filamentary ReRAM**
- Oxygen vacancy migration
- Thermo-Chemical Fuse/antifuse
- Electro-Chemical ECM

**Interfacial ReRAM**
- Cation Source (Ag⁺, Cu⁺ or...)
- Schottky or Tunnel Barrier
- Phase Change PCM

**Bulk Transition**
- Tunnel Magnetoresistance
- Electronic MIT (Mott)

**Switching Mechanisms**
- Filamentary ReRAM
- Interfacial ReRAM
- Bulk Transition

**Metallic Filament**
- Exchange layer
- Metal oxide
- Oxygen vacancy...

**Poly-crystalline**
- Amorphous
- Perovskite

**BIPOLAR**
- TMO: HfO₂, TaO₂
- CBRAM: Cu, Ag based
- Memristor VMCO, PCMO, TiO₂

**UNIPOLAR**
- TMO: NiO₂
- Chalcogenide alloys: GST
- STT RAM (CoFeB, MgO)

**Source:** M. Jurczak, imec, ISSCC 2015 Memory Forum
What is 3D XP?

3D XP is **NOT** ST-MRAM because:
- Can't build ST-MRAM in a XP array (1T-1R)
- Can't achieve 128Gb with ST-MRAM (30F²)
- ST-MRAM more expensive than DRAM

3D XP is **probably NOT** ReRAM because:
- ReRAM Latency is too slow for application
- ReRAM endurance may be too short for the application
- ReRAM is much more likely New Memory B (Cost focused)

Source: Intuitive Cognition Consulting estimates
Crosspoint resistive memory

(adapted from Burr, EIPBN 2008)

Source: R. Shenoy, IBM, IMW May 2013

PCMS

Unity Semiconductor 2 layer Conductive Metal Oxide RRAM

Ref: D. Eggleston, Flash Memory Summit 2011

Effective cell size: $4F^2$

Ref: D. Kau, IEDM 2009

Stack ‘L’ layers in 3D

Effective cell size: $4F^2/L$

Ref: T-Y. Liu, ISSCC 2013

3D multilayer stacked crosspoint arrays → path to low cost memory

WRONG!
XP Arrays: The Litho Problem

- XP array with shared layers requires multiple masking steps:
  - N+1 masking steps to form WL/BL
  - i.e. 4 XP layers requires a minimum of 5 masking steps
- XP cell size 4F² determined by WL/BL lithography:
  - To achieve sub 20nm F requires triple or quad critical layer patterning
  - i.e. 4 XP layers → 5 critical masking steps → x4 quad patterning → 20 times through the stepper at sub 20nm!!!

XP arrays will never be the lowest cost memory solution!

Source: Intuitive Cognition Consulting estimates

- N = 4 XP layers
- N+1 = 5 masking steps
- Cell size of 4F²
- Effective cell size 1F²
Reuse of the 3D NAND architecture and processing is the way forward for lowest cost RRAM.

Current Major Costs of Planar CMOx™:
- Multiple immersion patterning or EUV lithography
- Fine line interconnect very challenging

Source: C. Chevallier, Unity, CEATEC 2011
BUT, to do so requires a built-in selection capability. This is DIFFICULT!
See you in 2021!

Vertical ReRAM (VRRAM)

- No. of critical masks is not increasing with added memory layers
  - Multiple layers are formed at the same time
- Challenging to integrate the selector – self-rectifying cell needed

Reference:
S. Kim, LETI Memory Workshop 2012
I. G. Baek et al., IEDM 2011
H. S. Yoon et al., VLSI 2009
But **WHAT** 3D XP is doesn’t really matter…

…**HOW** will 3D XP be used?
How 3D XP will be used


Purley: Biggest Platform Advancement Since Nehalem

**PERFORMANCE FOR RANGE OF WORKLOADS**
- Better performance/Watt
- 1.5X memory bandwidth
- AVX-512 new instructions

**ALL NEW MEMORY ARCHITECTURE**
- Up To 4X capacity & lower cost than DRAM
- 500X faster than NAND
- Persistent data

**INTEGRATED NETWORK / FABRIC**
- 1/10G Ethernet
- 100G OmniPath™

**OPTIONAL INTEGRATED ACCELERATORS**
- QuickAssist™ encryption and compression offloads
- Skylake + FPGA
- Cannonlake graphics & media transcoding

Memory bandwidth increase and native support of persistent memory in Purley server platform.
How 3D XP will be used

2S Purley Platform Configuration Example

**Platform Ingredient Options**
- Fabric: Storm Lake Integrated, Storm Lake PCIe* card
- Storage: Downieville, SATA: Youngsville, PCIe*: Coldstream, PCIe* Cliffdale
- Software: Analytics, Efficiency, Performance, Secure Access, Tools
- Networking: Lewisburg as 4x10Gbe Integrated Network Solution with PHY, Fortville 4x10/2x40GbE (Controller), Sageville 2x10GBASE-T (Controller), Springville (1x1GBASE-T), Powerville (4x1GBASE-T)
  - Core Ingredients: Copperville 10GBASE-T, Jacksonville GbE PHY
- Accelerators: Intel® QuickAssist Accelerator Technology with Compression and Encryption, Lewisburg can also be used as PCIe* add-in card in end point mode
- Intel® Xeon Phi™ Product Family: Knights Corner/Landing Coprocessors and Processors
- FW/Bios: Manageability, Node Manager, Intel® RSTe, ME11, Intel® Trusted Execution Technology and Intel® Platform Protection Technology with Boot Guard

**“Apache Pass” DDR4 DIMMs supported in Purley.**

**Source:** The Platform, “Intel Lets Slip Broadwell, Skylake Xeon Chip Specs”, May 2015
Intel projects use of PCM memory as “Far memory”; uses DRAM as a “Near memory cache”. Consolidation of system memory, storage BIOS and TPM in PCM.

Intel PCM-Based DIMMs

Transactional protocol over DDR bus handles non-deterministic behavior of PCM.

Intel PCM-Based DIMM Controller

Writes are persistently buffered in the PCM controller to mask write latency.

**FIG. 5E**

Speculation on Intel usage of 3D XP

- **Server memory DIMMs:**
  - In-memory compute, Real time analytics applications
  - Coincide with Purley platform launch in 2017
  - DDR4 bus with transactional protocol (non-JEDEC)
  - Read is PCM native latency (50-200ns)
  - Write is buffered in PCM controller (latency hidden)
  - 100’s of GBs per DIMM (3x-4x DRAM capacity/DIMM)
  - Trouble meeting 12W power window (PCM energy)

Source: Intuitive Cognition Consulting estimates
Strengths

- Brand new memory type brings new capabilities
- Semi-Persistent memory arrives!
- Server memory target competes with DRAM cost
- Read/write performance near full DRAM speed
- Full system approach

Weaknesses

- 3D XP cost will be high:
  - DRAM $8/GB, 3D XP $4/GB, 3D NAND $0.2/GB
- Sole sourced from Intel/Micron
- Requires major hardware and software changes
- >12W Power requires server thermal re-design
- Non-JEDEC standardized DDR4 transactional interface

Source: Intuitive Cognition Consulting estimates
Making Sense of 3D XP

• **WHY** are you surprised?
• **WHAT** is it (really)?
• XP Array Limitations: **The Litho Problem**
• **HOW** will it be used?
• 3D XP **strengths & weaknesses**