BOISE STATE UNIVERSITY

Wide I/O DRAM Architecture Utilizing Proximity Communication

by

Qawi Harvard

Thesis Defense – October 8th, 2009

College of Engineering

Introduction

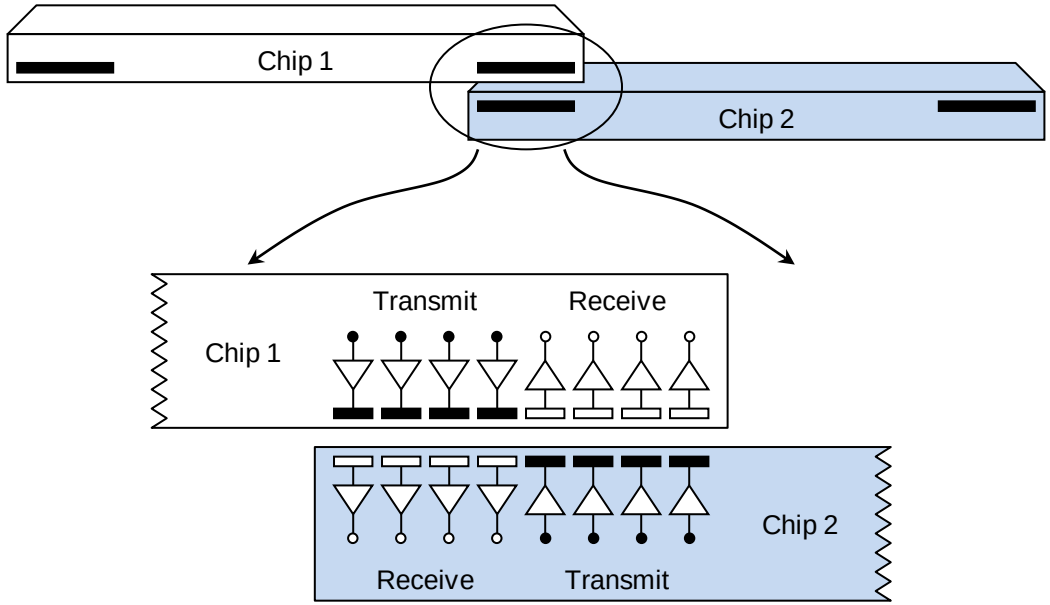
Bandwidth and power consumption of dynamic random access memory stifles computer performance scaling

- Background
- Status of Proximity Communication
- DRAM Market Analysis
- 4 Gb DRAM Architecture
- Wide I/O DRAM Architecture Utilizing Proximity Communication

Background

- Memory Gap
  - ✓ Main memory does not scale with processor performance
- Power
  - ✓ Current consumption is rising
  - ✓ Bandwidth increases power
  - ✓ Voltage scaling masks the issue
- Density
  - ✓ Memory channel loading
  - ✓ Limits bandwidth
- Proximity Communication
  - ✓ Proposed by Ivan Sutherland – US Patent #6,500,696
  - ✓ Promises to reduce power and increase bandwidth

Proximity Communication



- Capacitive Coupled Proximity Communication
  - ✓ Top metal forms the parallel plates
  - ✓ Chip-to-chip communication through coupling capacitor

Ref:[1]

Proximity Communication

- Benefits
  - ✓ Increased I/O density
  - ✓ Avoids on/off chip wires
  - ✓ Eases chip replacement at the system level
  - ✓ Enhances system level testability
  - ✓ Enables smaller chip sizes
  - ✓ Removes the need for ESD protection
- Challenges
  - ✓ Mechanical misalignment
  - ✓ Applying power to the chips
  - ✓ Thermal solution

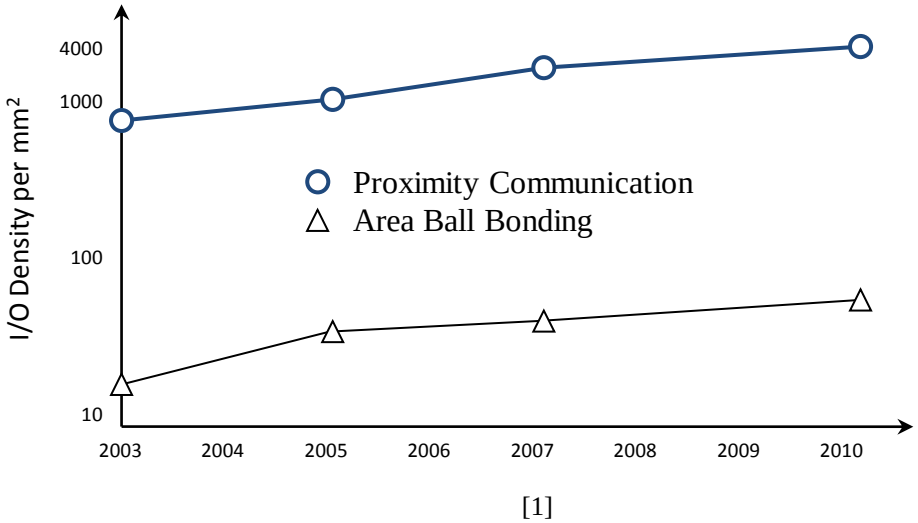
Ref:[1-5]

Proximity Communication

- Parallel Plate Capacitance

$$C = \frac{\varepsilon_0 A}{d} \qquad \varepsilon_0 = 8.9\ {}^{aF}/_{\mu m}$$

- 10 pF/mm^2
  - ✓ Chip-to-chip separation
  - ✓ d = 1 µm
- One channel
  - ✓ 50 fF
  - ✓ 200 signals/mm^2

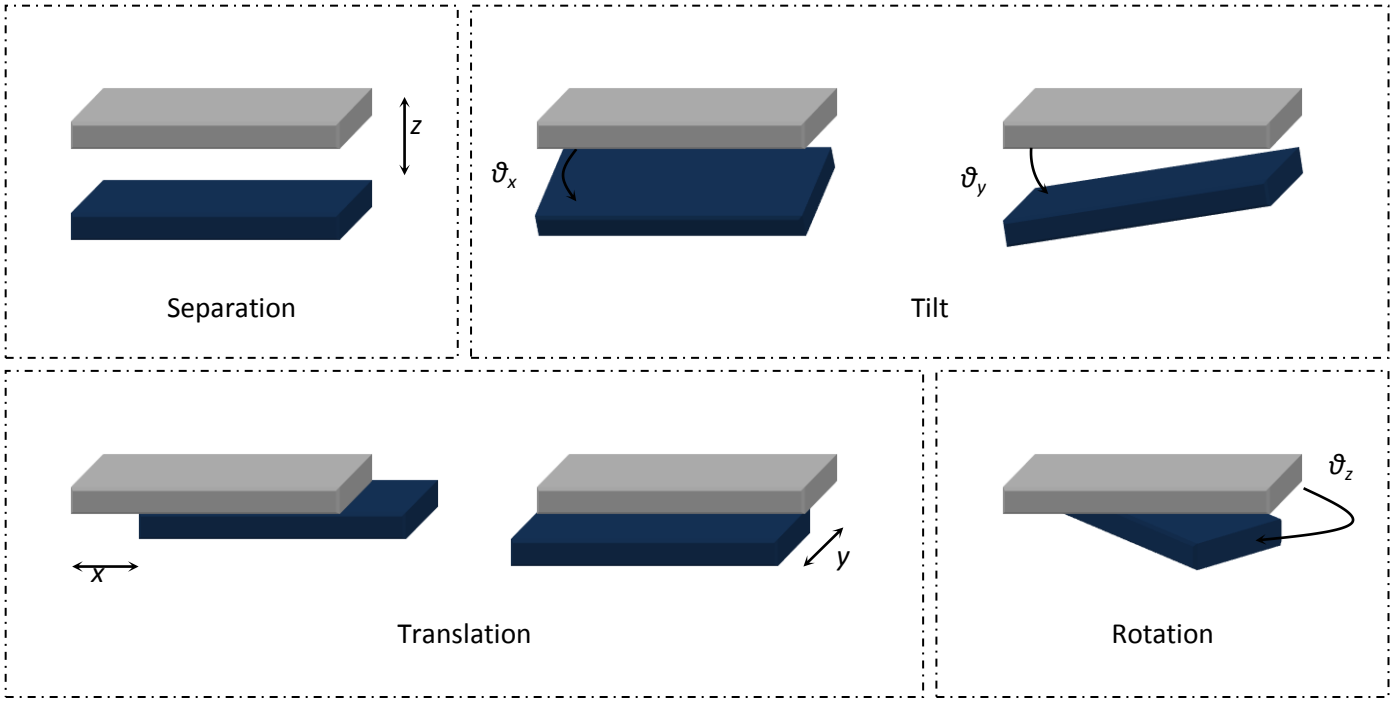


[1]

Ref:[1]

Proximity Communication

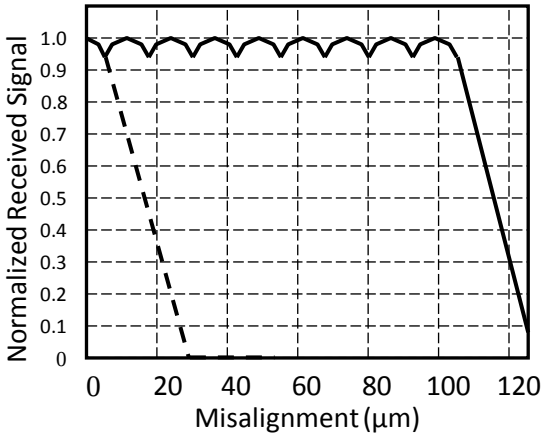
- Mechanical Misalignment
  - ✓ Six axis
  - ✓ Multiple sources

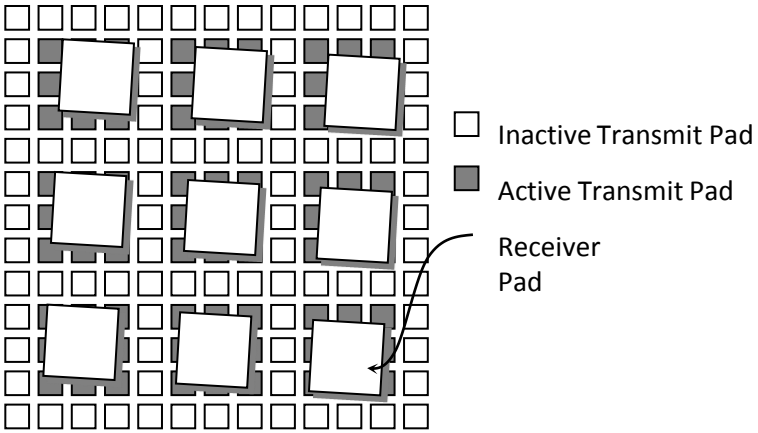


Ref:[5]

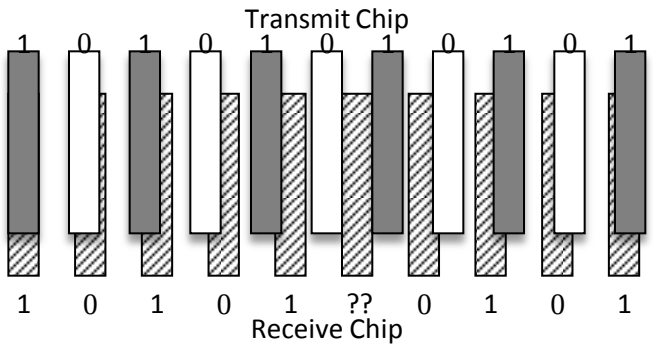
Proximity Communication

- Electronic Sensors
  - ✓ Chip-to-chip separation sensors (0.2 µm resolution)
  - ✓ Vernier scale incorporated on chip (1.0 µm resolution)
- Electrical Re-Alignment
  - ✓ Receive array
  - ✓ Micro-transmit array
  - ✓ Electronic steering circuit



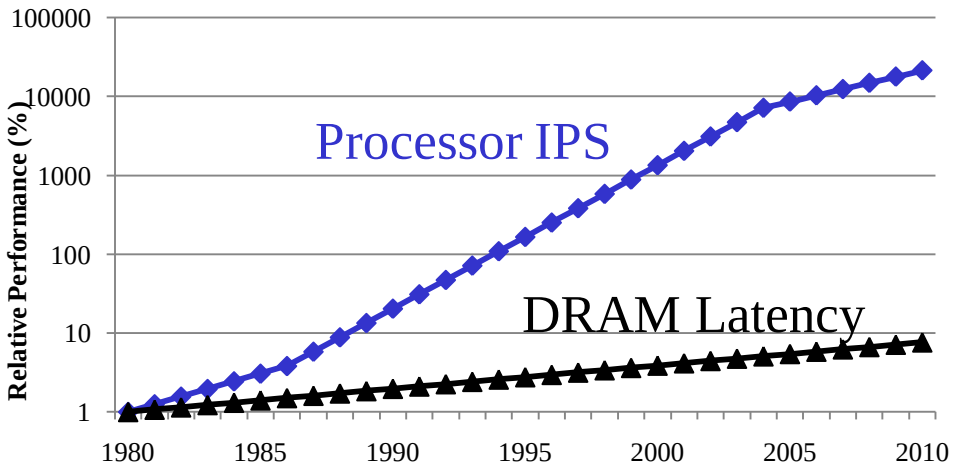


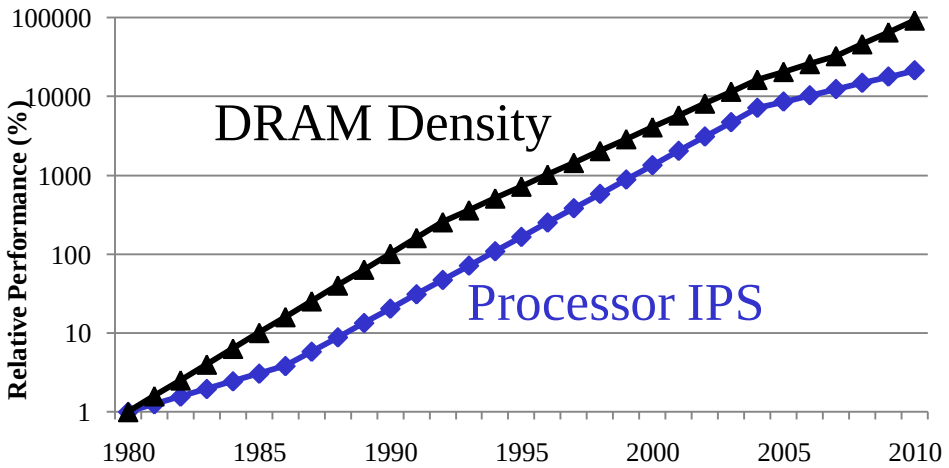
Ref:[1-5]



DRAM Market Analysis

- Revisit the Memory Gap
  - ✓ "Performance" becomes a relative term
  - ✓ Dichotomy in scaling
- Why Density?
- Why Not Latency?

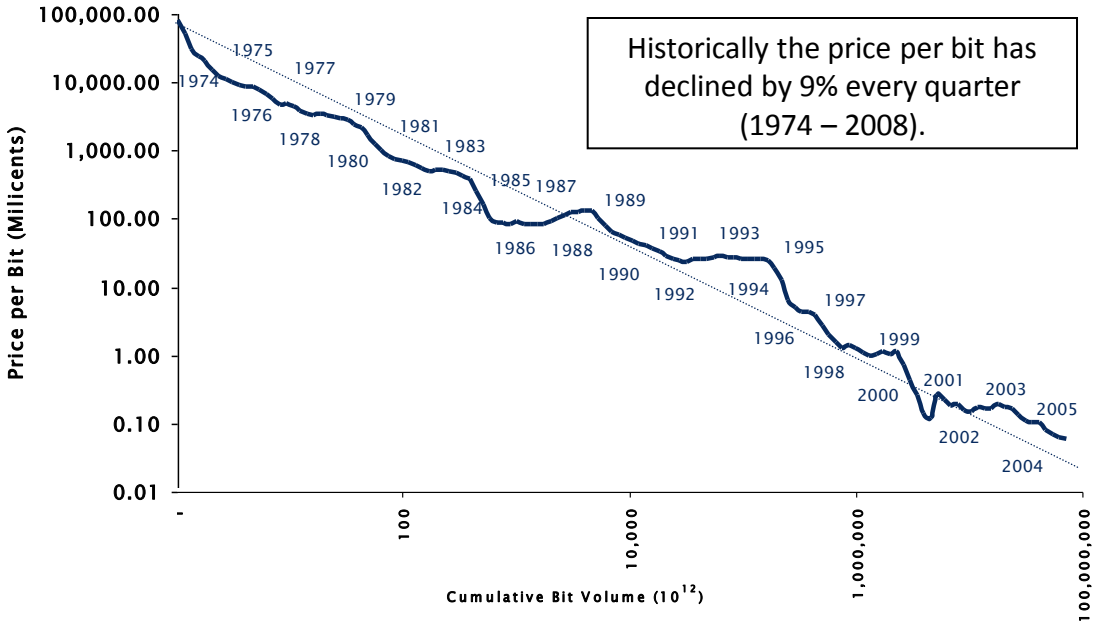




Ref:[6]

DRAM Market Analysis

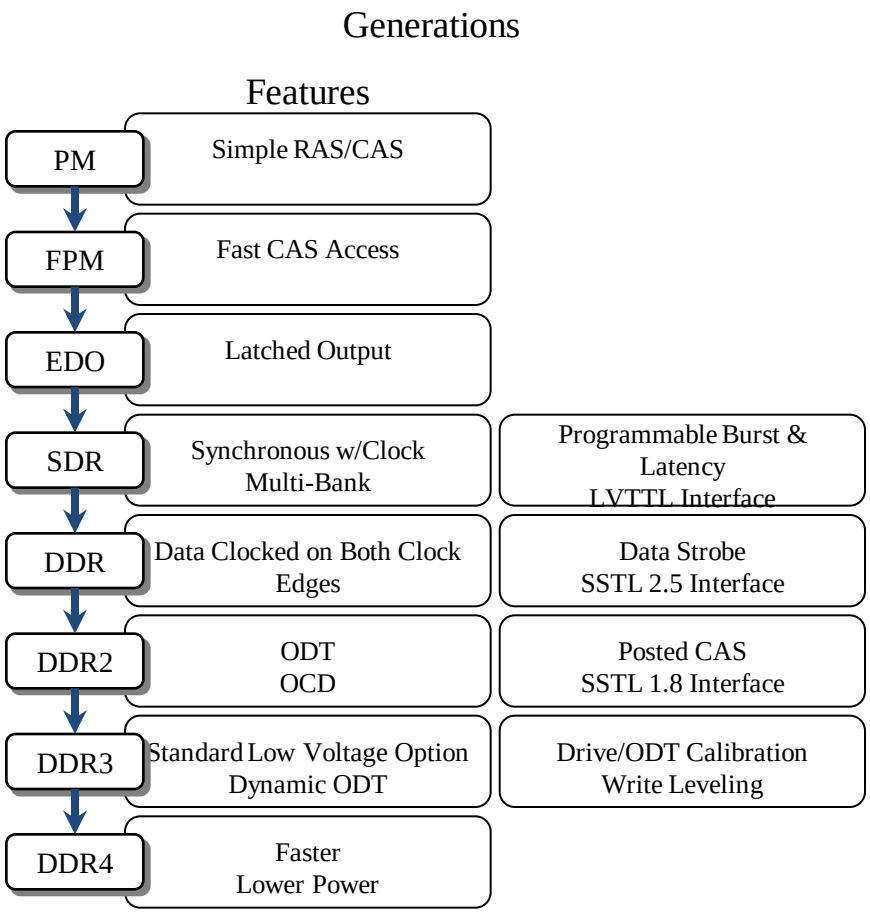
- Moore's Law
  - ✓ 41% increase in transistor count per year
- Selling Price
  - ✓ 36% historic decline per year
- Putting it into Perspective
  - ✓ 1 Gb 2009 → $2.00
  - ✓ 2 Gb 2011 → $1.64
- Density or Bust!!



Ref:[7-8]

DRAM Market Analysis

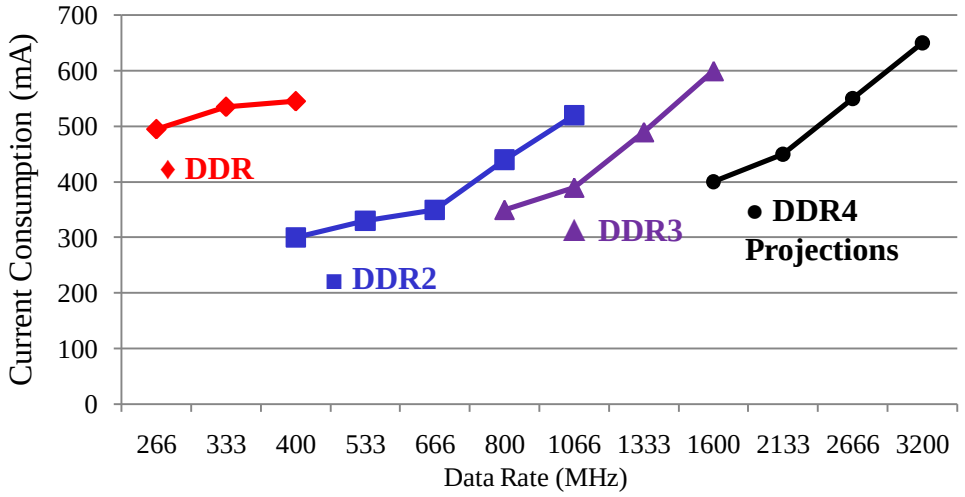
- Cost
  - ✓ Low cost manufacturing process
    - o 3 metal layers
      - Increased usage of each level
    - o Small chip size
      - Limits I/O count
- Moore's Law
  - ✓ 41% scaling per year
    - o Wordline cross sectional area
    - o Tight metal pitch
    - o Contact resistance
- Physics of Scaling
  - ✓ Latency must increase

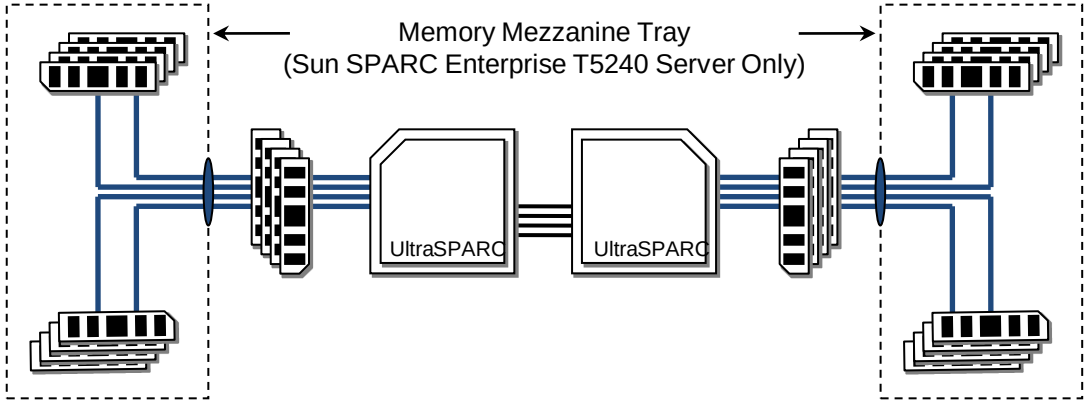


Ref:[7-8,13]

DRAM Market Analysis

- Home PC
  - ✓ Plugged into a wall
- Mobile
  - ✓ Battery life
- Server
  - ✓ Power consumption
  - ✓ Cooling
- Trending Up
  - ✓ Poor Efficiency
  - ✓ Bandwidth Driven

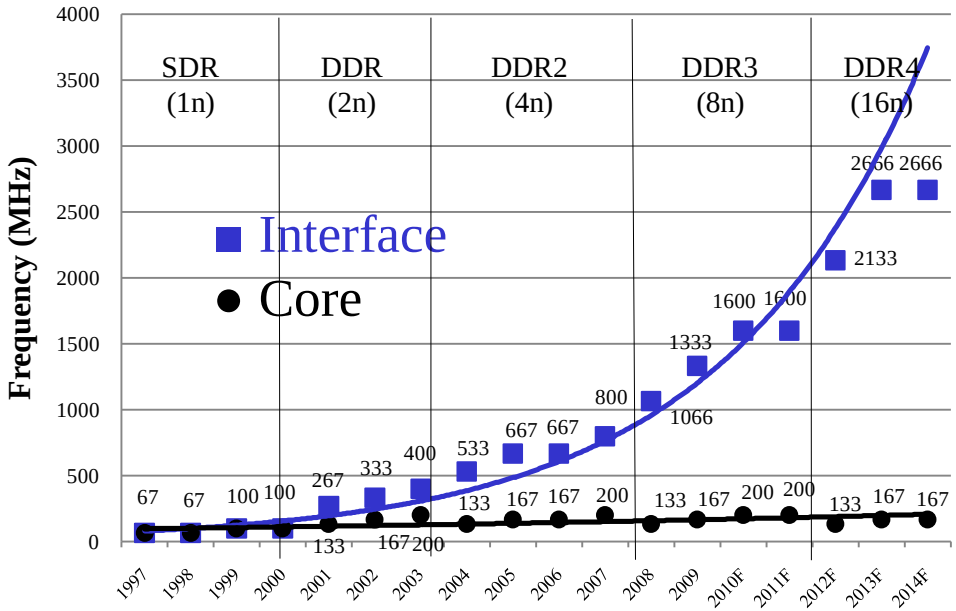


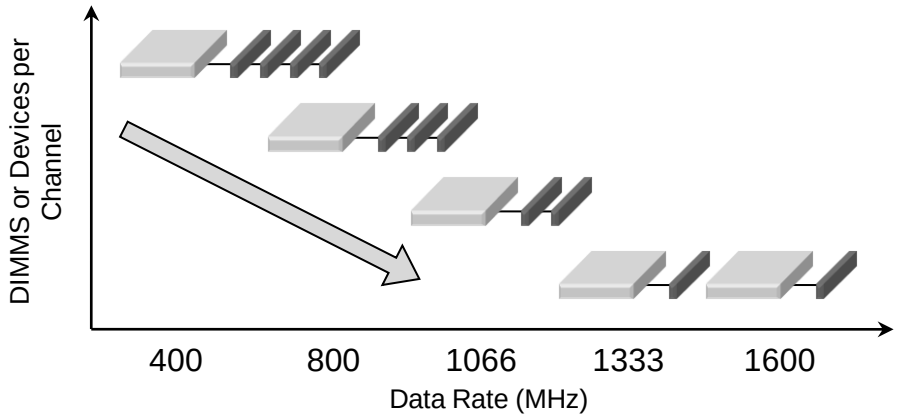


Ref:[14-17]

DRAM Market Analysis

- Interface versus Core
  - ✓ Interface bares the burden
  - ✓ Core cycles
- DRAM Pre-fetch
  - ✓ Doubles at each generation
- Density limited by bandwidth
  - ✓ SSTL loading in memory channel
  - ✓ Increase chip count per module

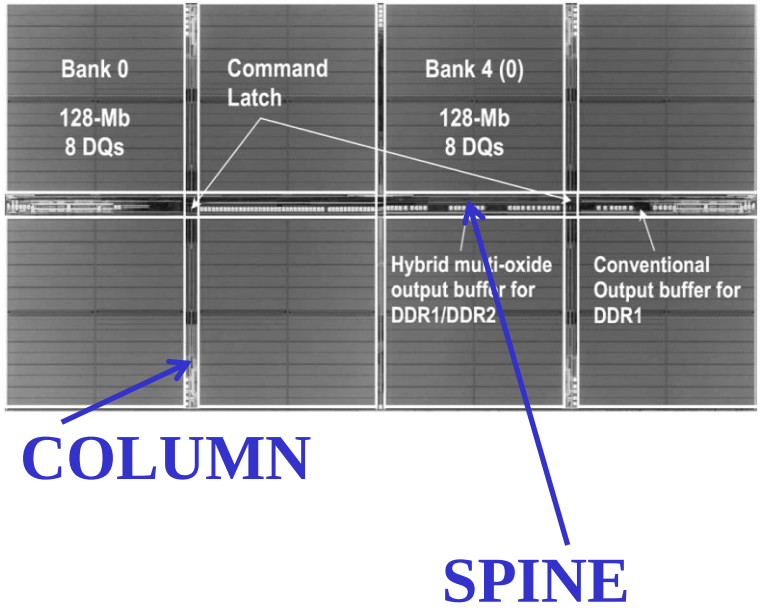


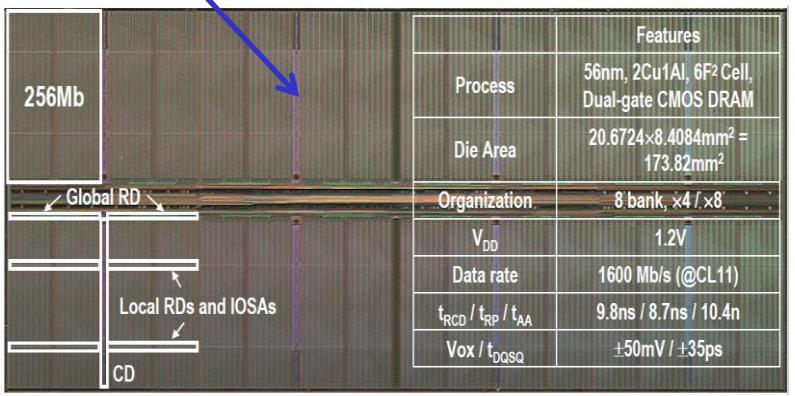


Ref:[14-17]

4 Gb DRAM Architecture

- Possible Architecture
  - ✓ Compared to ITRS
- 2012 Production Release
- 74 mm^2
- 56 % Array Efficiency
- 40 nm

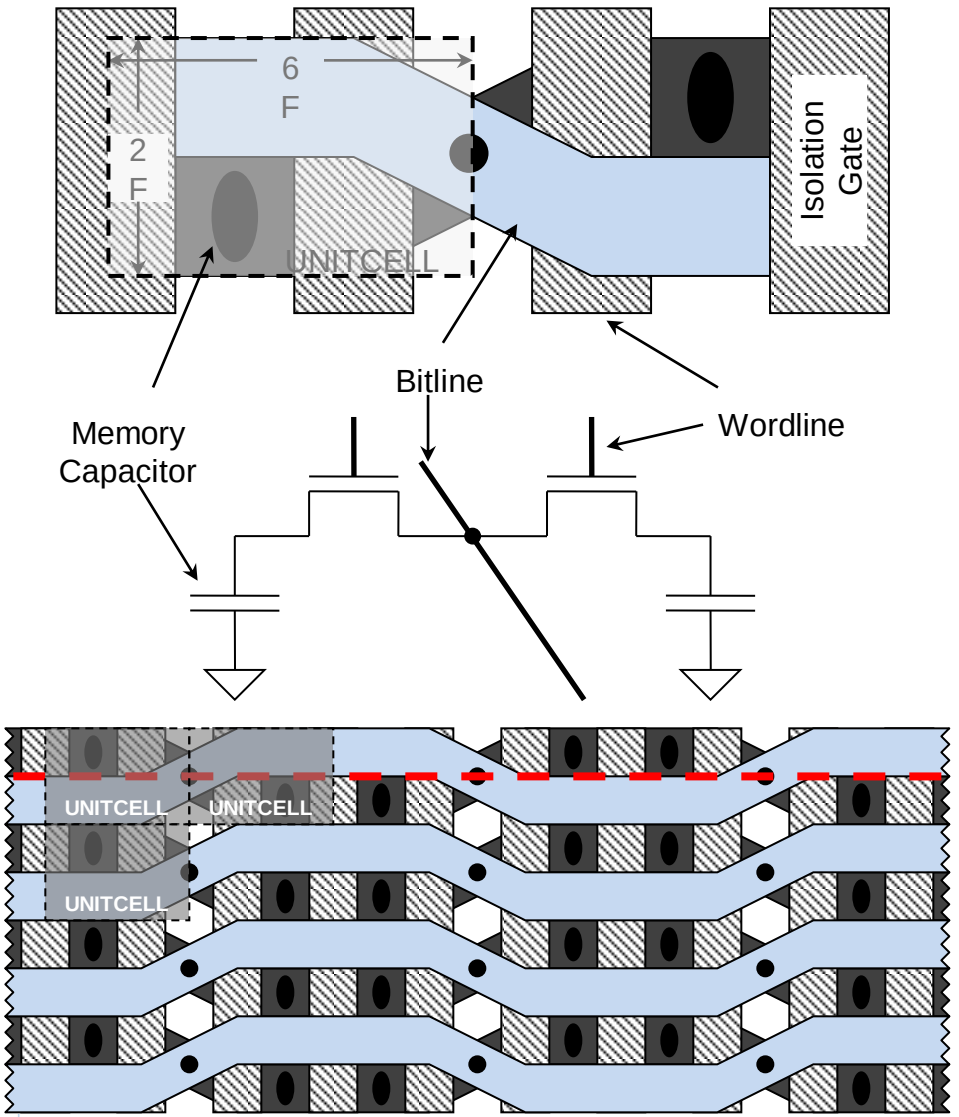




Ref:[10,12,20-24]

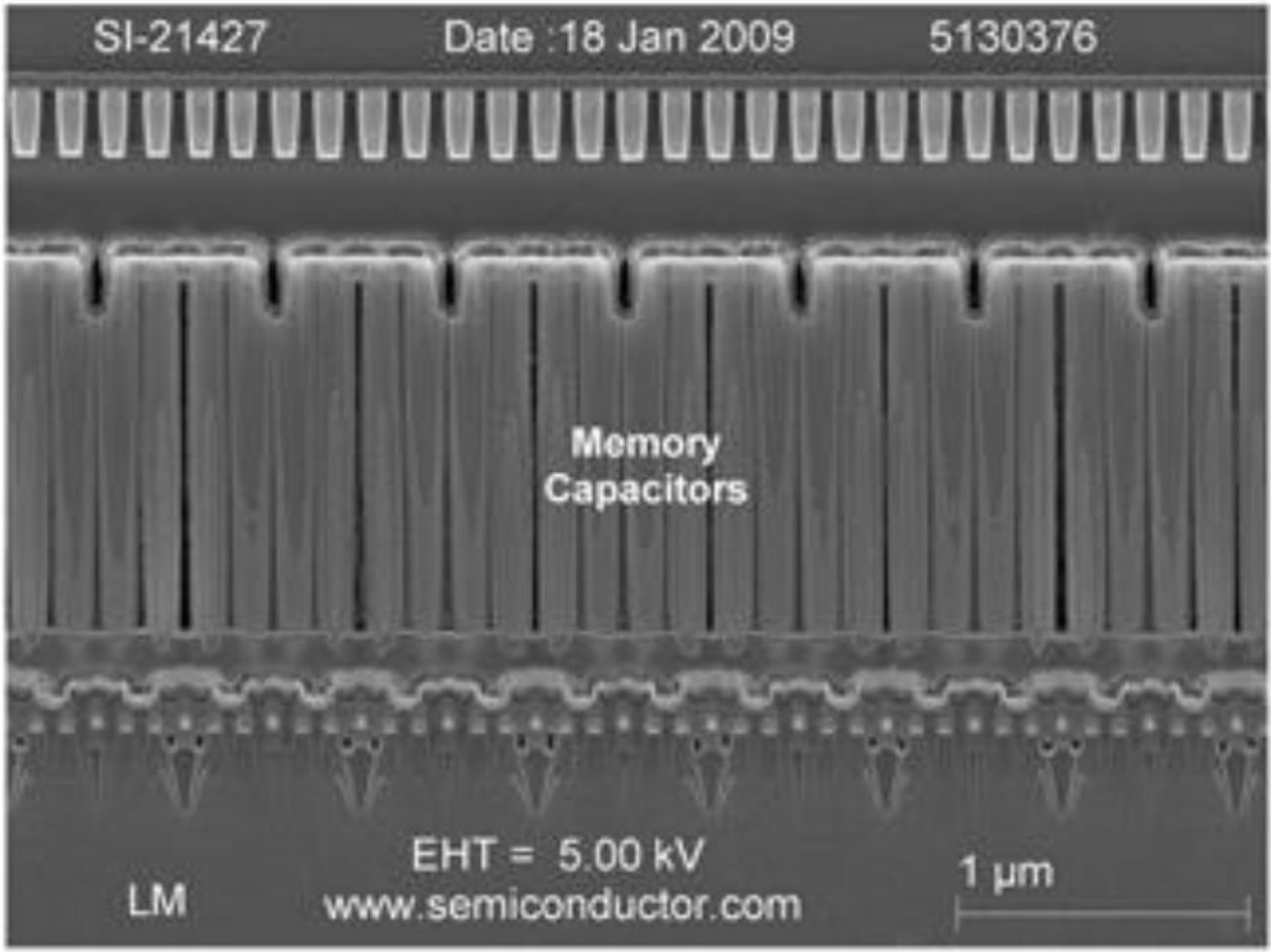
4 Gb DRAM Architecture

- $6F^2$ Memory Cell
- Feature Size = 40 nm
- Cell Area = 0.0096 μm^2
- 3F Pitch per Wordline
- 2F Pitch per Bitline
- 256 kb Array Macro
  - ✓ Core Array
  - ✓ 512 bitlines ≈ 43.6 µm
  - ✓ 512 wordlines ≈ 65.4 µm
- Periphery Circuitry
  - ✓ 4 µm space allocated



Ref:[24]

4 Gb DRAM Architecture



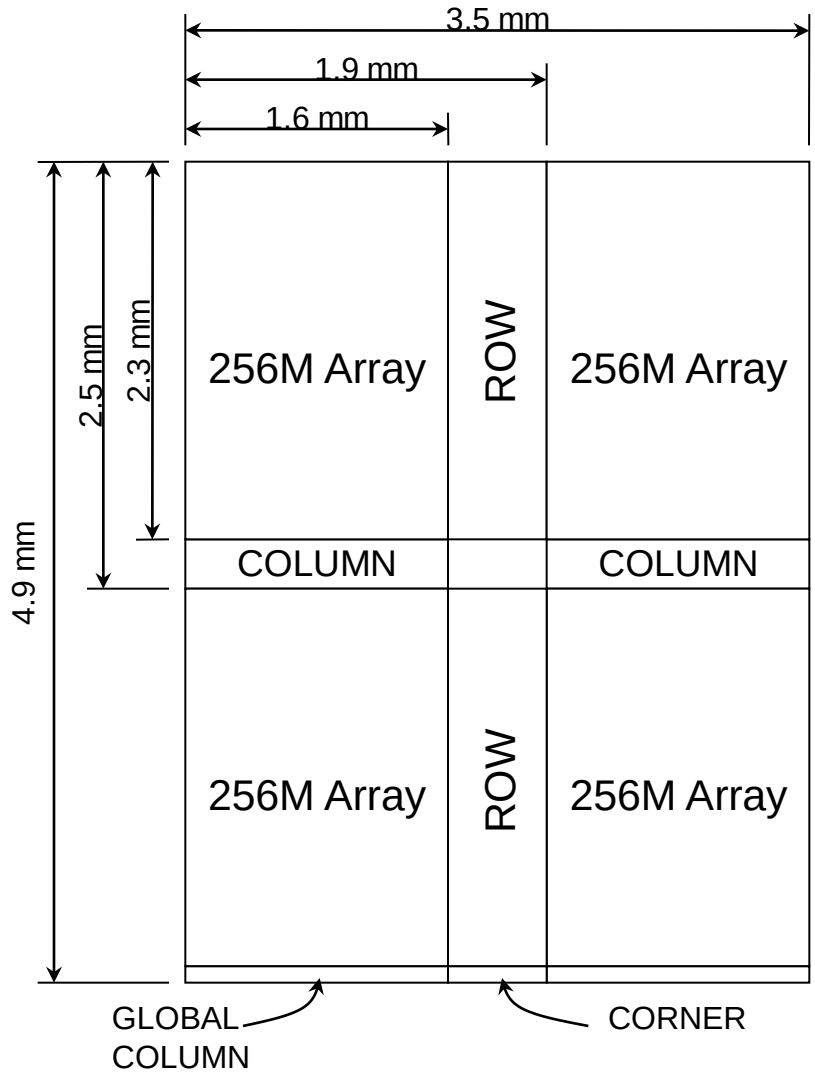
Ref:[25]

4 Gb DRAM Architecture

- 256 Mb Array
  - ✓ 32 x 32 256 kb macros

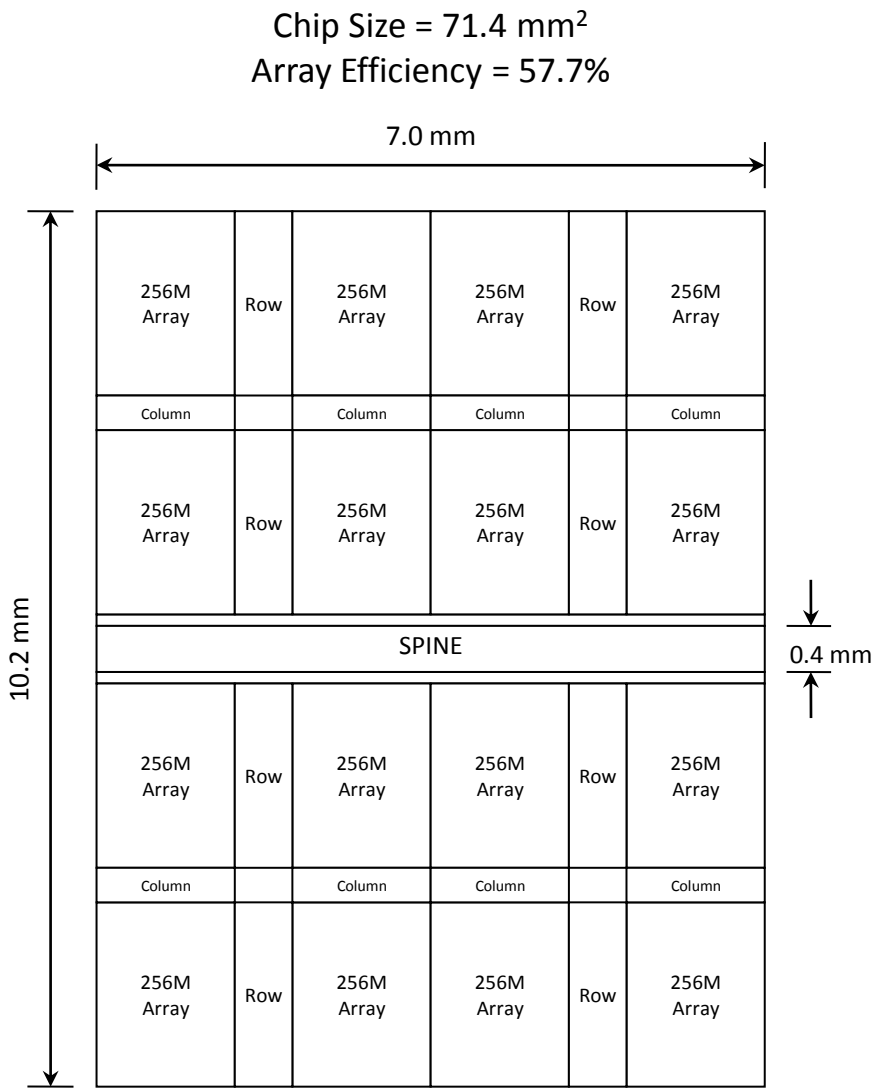
$$32 \cdot \left(43.6\,\mu m + 4\,\mu m\right) = 1.532\ mm$$

- 1 Gb Array
  - ✓ Multiple implementations



4 Gb DRAM Architecture

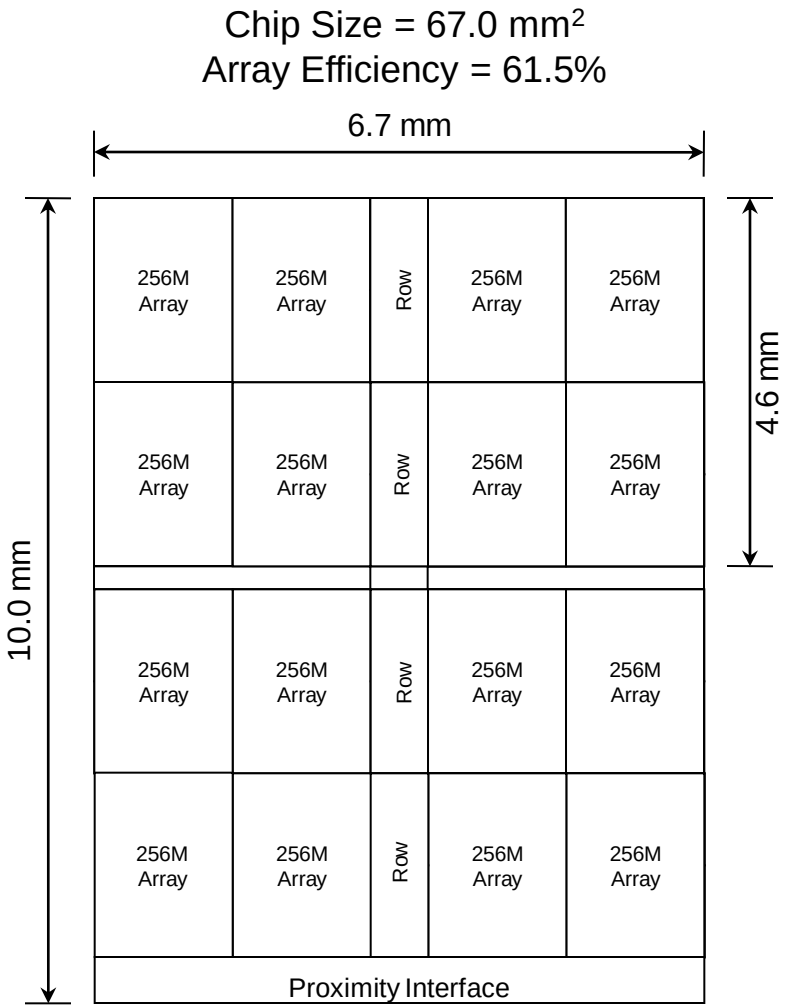
- ITRS
  - ✓ 74 mm^2
  - ✓ 56% Array Efficiency
- Wide I/O Architecture
  - ✓ Moving the pads
  - ✓ Centralized Row
  - ✓ Centralized Column



Ref:[29]

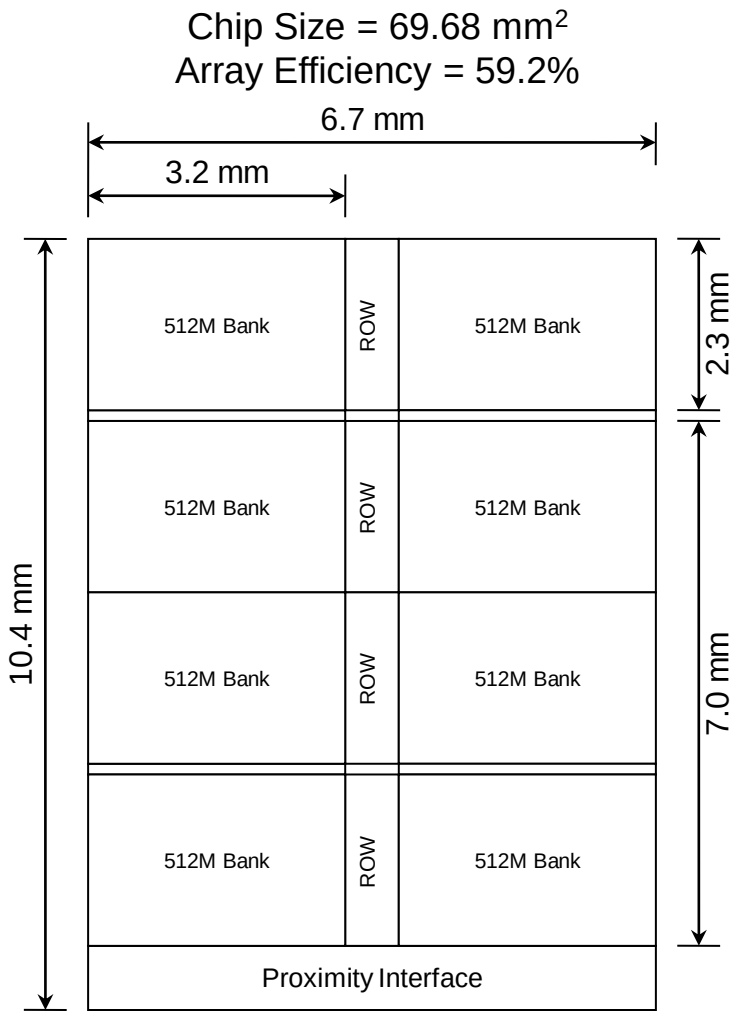
Wide I/O Chip Architecture

- 64 Bytes per Chip
- 6.2% Chip Size Reduction
- 6.6% Increase in Array Efficiency
- Challenges
  - ✓ Routing from the edge
  - ✓ Array I/O route increase
    - o 2.3 mm → 4.6 mm
  - ✓ Additional row decode
- Create Eight Internal Banks



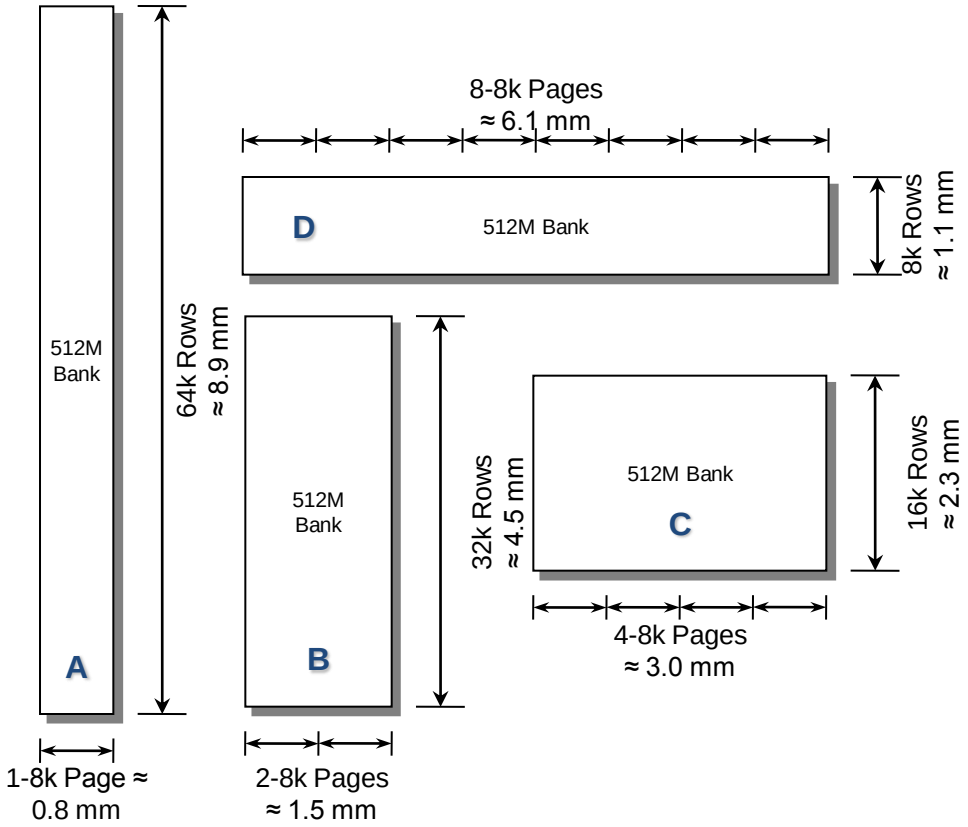
4 Gb DRAM Architecture

- 1 mm Allocation for Proximity Channel
- Buffers at the Center
  - ✓ Increase global I/O metal usage
- Array I/O Routing Reduced to 2.3 mm
- Architecture NOT Efficient for Proximity Communication
  - ✓ 6.7 mm versus 10.4 mm
  - ✓ Buffers required
  - ✓ Large metal usage



4 Gb DRAM Architecture

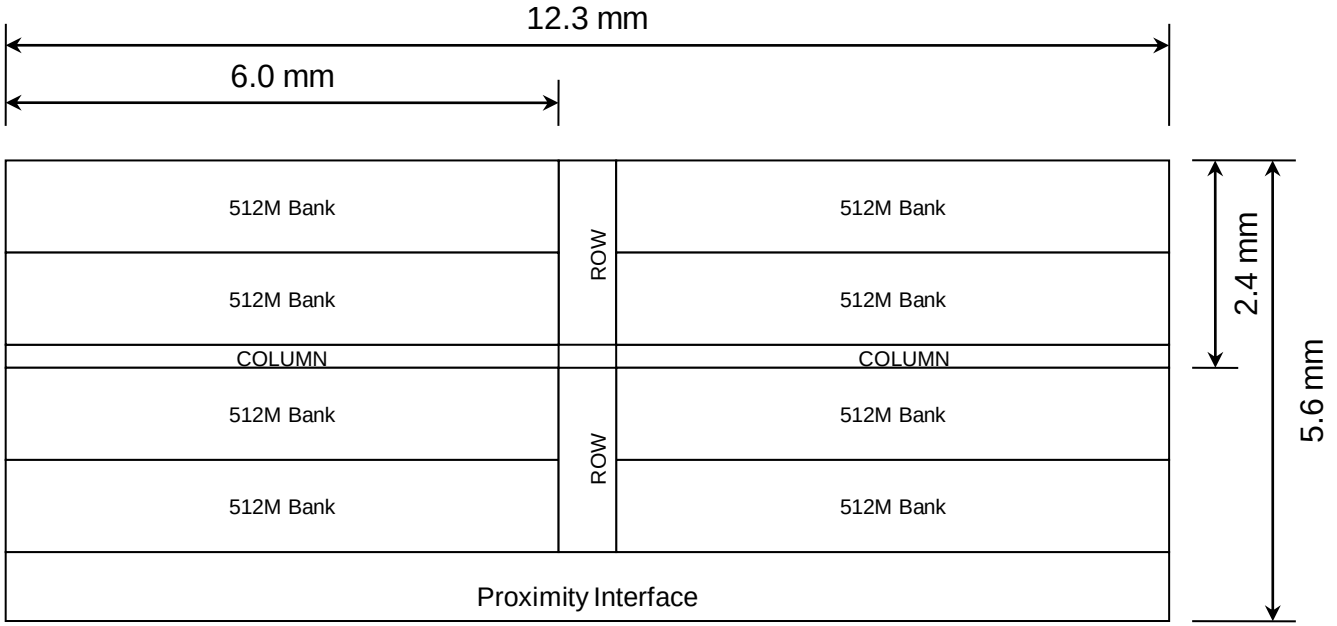
- Multiple Bank Architectures
- Page Size
  - ✓ Standard size = 8k
  - ✓ Energy efficiency
- Global Row Routing
  - ✓ ~20 ns latency
- Global Column Routing
  - ✓ ~5 ns latency
- D – Architecture
  - ✓ Page decode



Ref:[13,16,27,30,]

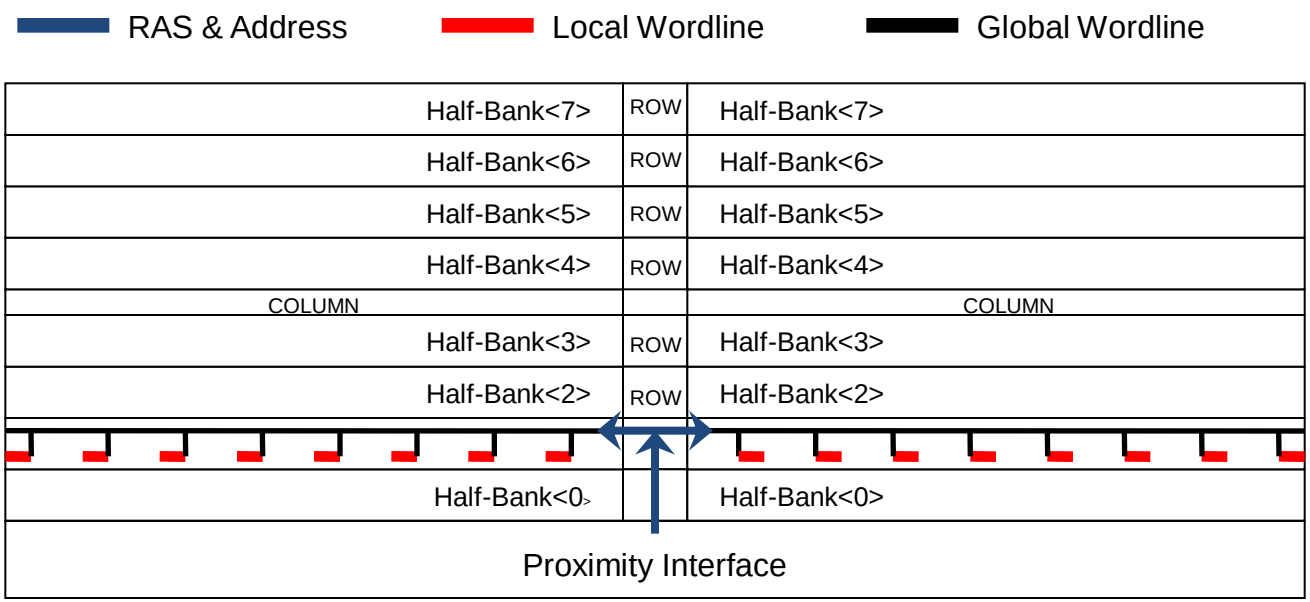
Wide I/O Chip Architecture

- Chip Size = 68.88 mm^2, Array Efficiency = 59.9%
  - ✓ Centralized row & column
  - ✓ Buffers not required
  - ✓ 12.3 mm for proximity communication
  - ✓ Enables two levels of metal



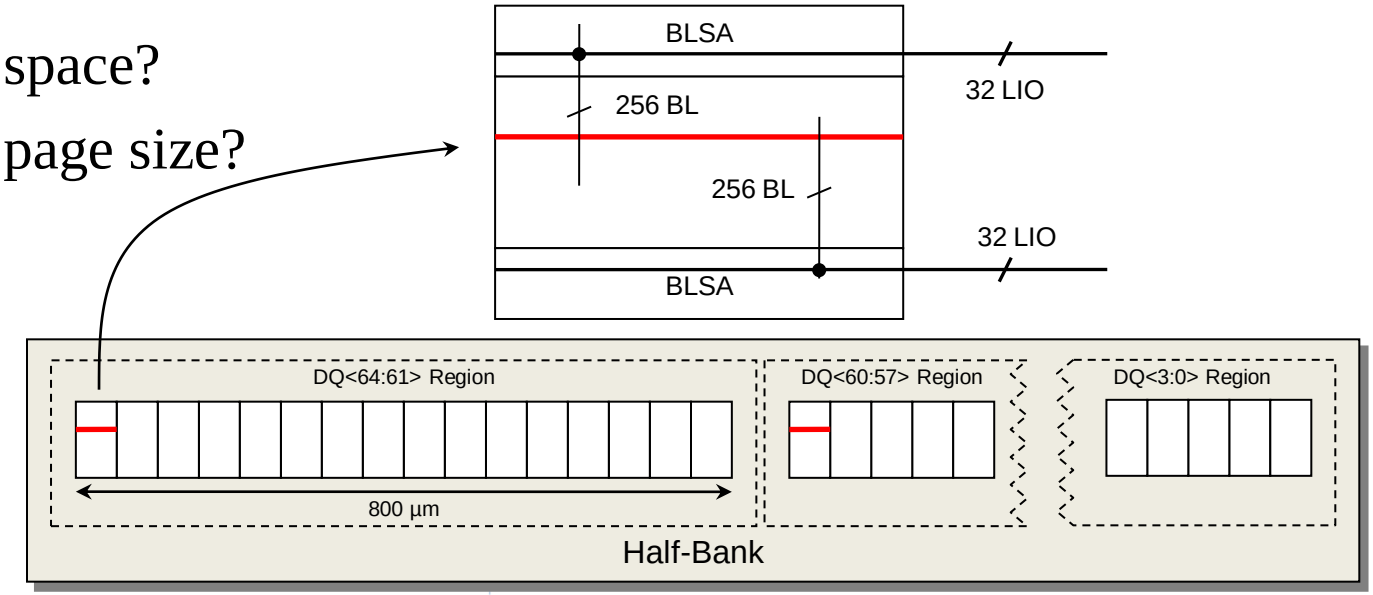
Wide I/O DRAM Architecture

- Split Bank Architecture
  - ✓ 64 bytes = 512 signals
  - ✓ 6 mm / 256 ≈ 43 µm per signal
  - ✓ 0.4 µm pitch < 1 % metal usage



Wide I/O DRAM Architecture

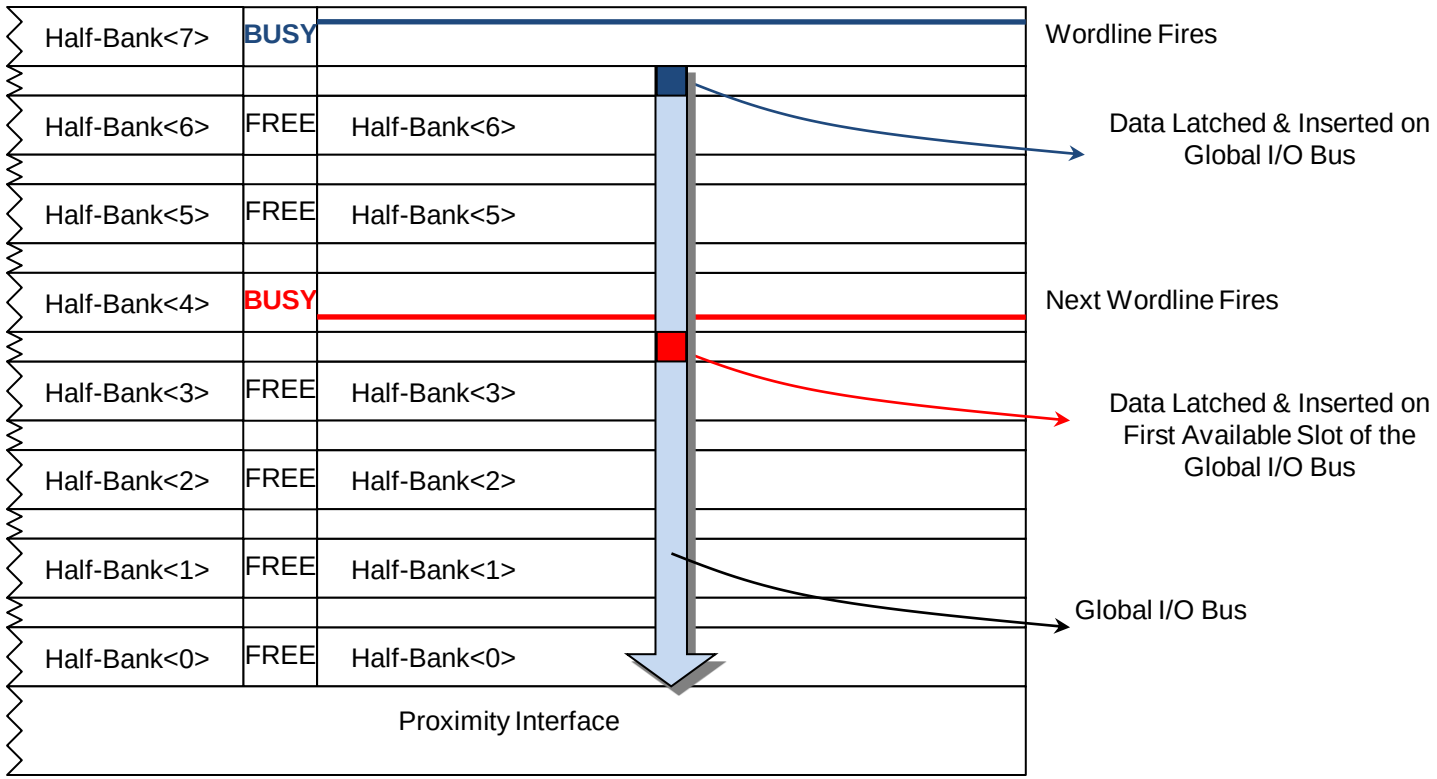
- Split Page Architecture
  - ✓ 8k page keeps current relative
  - ✓ Page decode required
  - ✓ 32 differential signals per macro
- Local I/O Routing
  - ✓ Space limited
  - ✓ Increase space?
  - ✓ Increase page size?



Ref:[12,20-24,26,28]

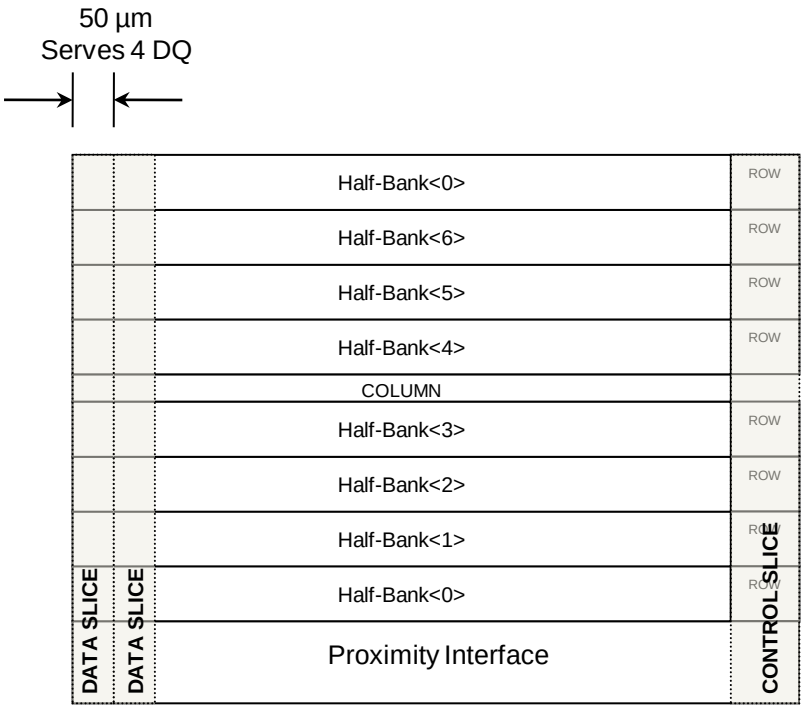
Wide I/O DRAM Architecture

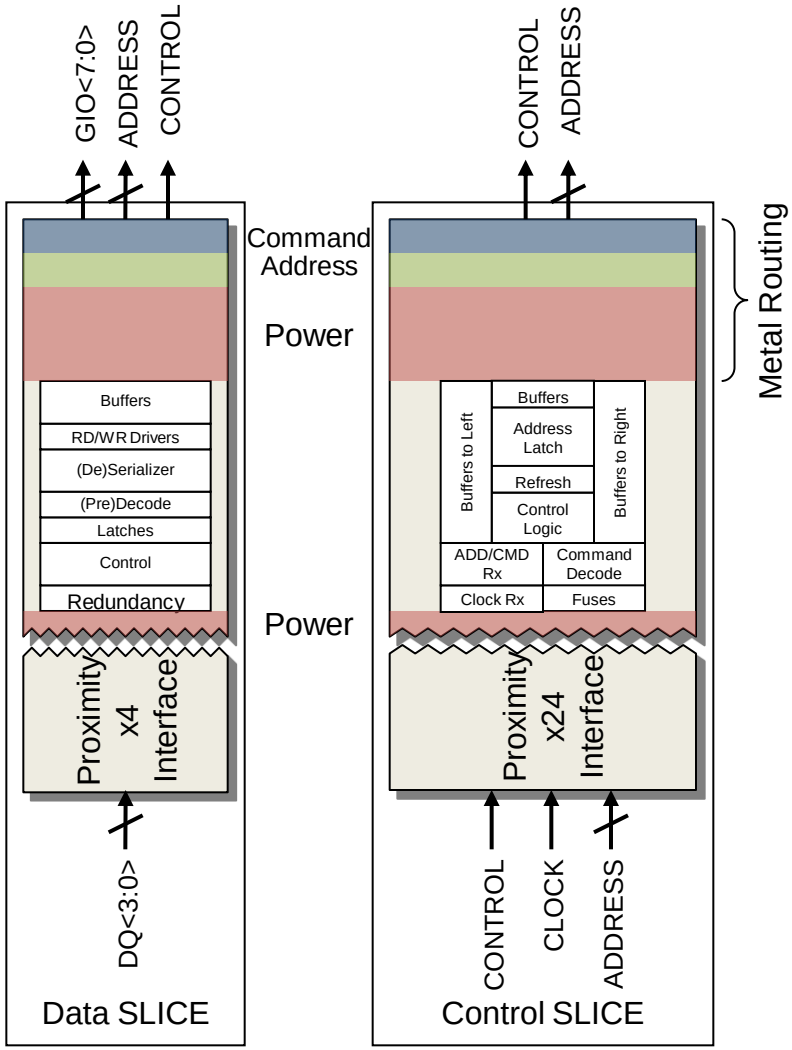
- New Column Routing
  - ✓ Global I/O operates at higher frequency
  - ✓ Protocol allows for insertion of data



Wide I/O DRAM Architecture

- Slice Architecture
  - ✓ Ease of design
    - o Uniformity, speed, verification



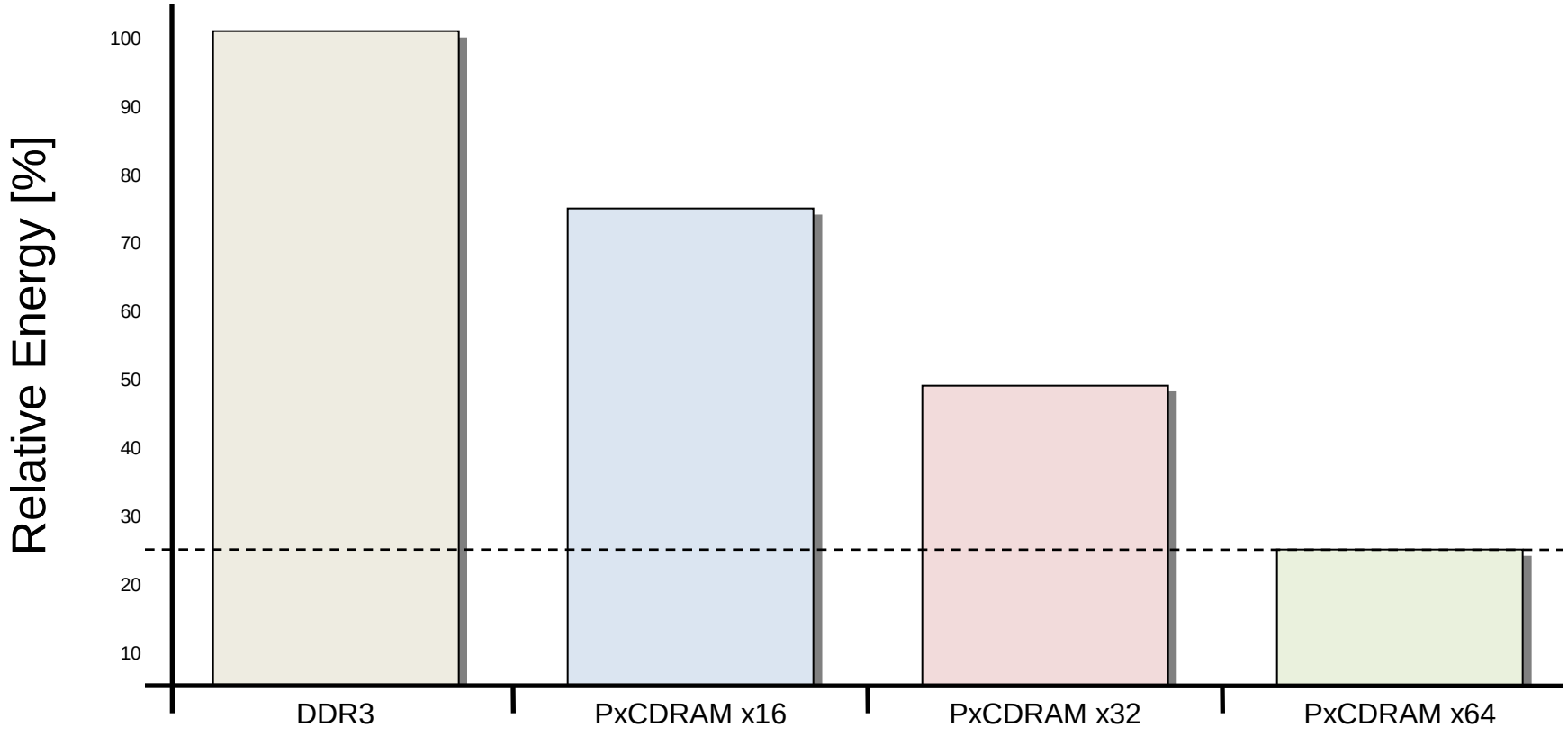


Wide I/O DRAM Architecture

- 64 Bytes per Chip
  - ✓ Significant bandwidth increase
- Power Consumption
  - ✓ Standard 8k page size
  - ✓ Split bank, split page
- Cost Performance
  - ✓ Two metals enabled for 4 Gb
  - ✓ Smaller chip size, higher array efficiency

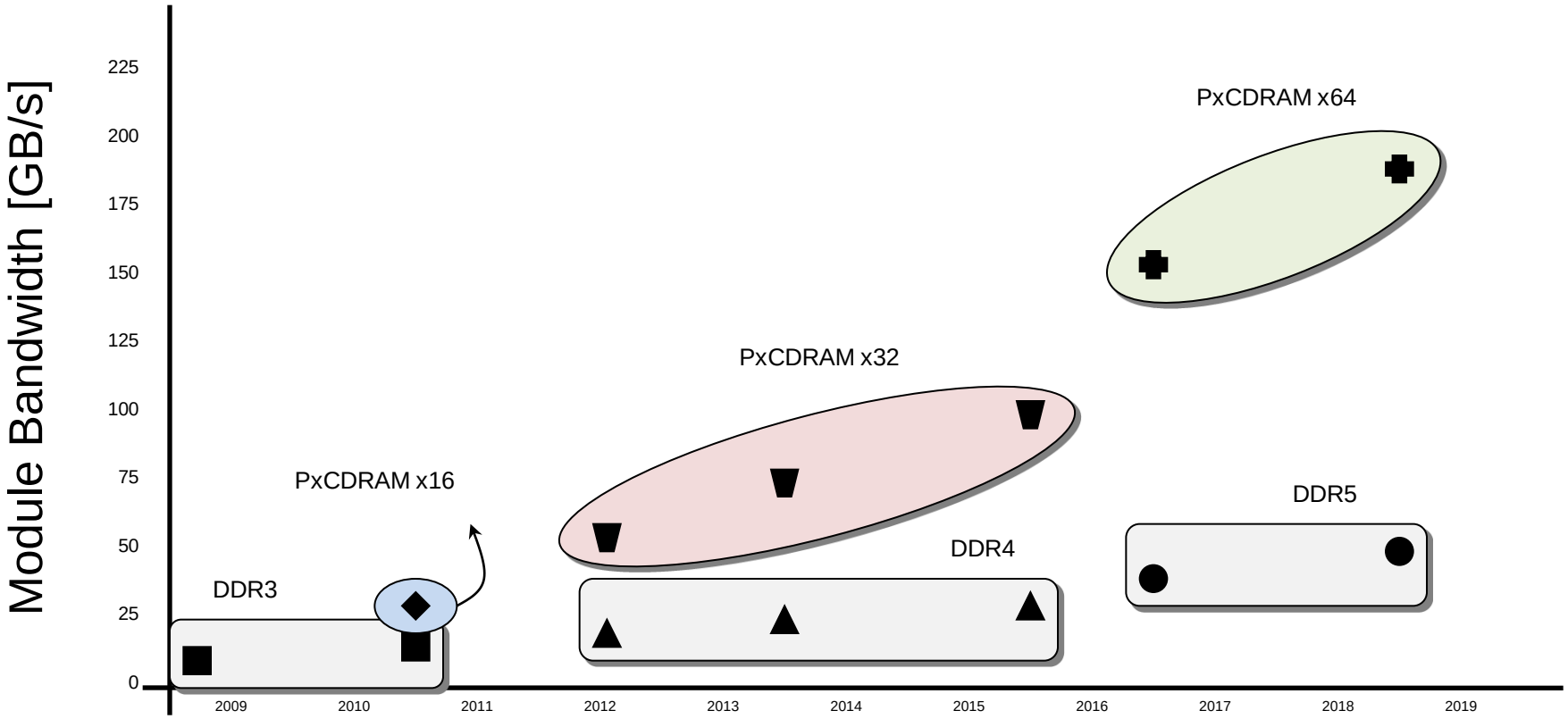
Wide I/O DRAM Architecture

- Power Consumption



Wide I/O DRAM Architecture

- Bandwidth



Future Work

- Applying Proximity Communication to New Memory Technologies
  - ✓ "High" density
  - ✓ Chalcogenide
  - ✓ Slice architecture
  - ✓ Circuit design techniques
- Local I/O Routing
  - ✓ New column global I/O structure
  - ✓ Through bitline routing
  - ✓ Novel local I/O latch

Acknowledgments

- Dr. Jake Baker
- Dr. Kris Campbell
- Dr. Robert Drost
- Dr. Sin Ming Loo
- Dr. Thad Welch
- Ms. Donna Welch
- Family support

- Questions?

References

[1] R. Drost, R. Hopkins, I. Sutherland, "Proximity Communication," *Proceedings of the IEEE 2003 Custom Integrated Circuits Conference*, vol. 39, issue 9, pp. 469-472, September 2003.

[2] D. Salzman, T. Knight, "Capacitively Coupled Multichip Modules," *Multichip Module Conference Proceedings*, pp. 487-494, April 1994.

[3] R. Drost, R. Ho, R. Hopkins, I. Sutherland, "Electronic Alignment for Proximity Communication," *IEEE International Solid State Circuits Conference*, vol. 1, pp. 144-145, February 2004.

[4] D. Hopkins, A. Chow, R. Bosnyak, J. Ebergen, S. Fairbanks, J. Gainsley, R. Ho, J. Lexau, F. Liu, T. Ono, J. Schauer, I. Sutherland, R. Drost, "Circuit Techniques to Enable 430Gb/s/mm^2 Proximity Communication," *IEEE International Solid State Circuits Conference*, pp. 368-369, pp. 609, February 2007.

[5] A. Chow, D. Hopkins, R. Ho, R. Drost, "Measuring 6D Chip Alignment in Multi-Chip Packages," *Proceedings of IEEE Sensors*, pp. 1307-1310, October 2007.

[6] J. Hennessy, D. Patterson, *Computer Architecture A Quantitative Approach*, *4th ed.*, Morgan Kaufmann Publishers, San Francisco, 2007. ISBN 978-0-12-370490-0

[7] G. Moore, "Cramming more components onto integrated circuits," *Electronics Magazine*, pp. 4-6, April 1965.

[8] D. Klein, "The Future of Memory and Storage: Closing the Gap," *Microsoft WinHEC 2007*, May 2007.

[9] B. Pang, Caris & Company http://www.semi.org/cms/groups/public/documents/web_content/p043628.pdf, March 2008.

[10] K. Kim, G. Jeong, "Memory Technologies for sub-40nm Node," *IEEE International Electron Device Meeting*, pp. 27-30, December 2007.

[11] J. Burnim, "On the Scaling of Electronic Charge-Storing Memory Down to the Size of Molecules," *The MITRE Corporation*, November 2001.

[12] Y. Park, S. Lee, J.W. Lee, J.Y. Lee, S. Han, E. Lee, S. Kim, J. Han, J. Sung, Y. Cho, J. Jun, D. Lee, K. Kim, D. Kim, S. Yang, B. Song, Y. Sung, H. Byun, W. Yang, K. Lee, S. Park, C. Hwang, T. Chung, W. Lee, "Fully Integrated 56 nm DRAM Technology for 1Gb DRAM," *IEEE Symposium on VLSI Technology*, pp. 190-191, June 2007.

[13] D. Rhosen, "The Evolution of DDR," *VIA Technology Forum*, 2005.

[14] SUN Microsystems, "SUN SPARC Enterprise T5120, T5220, T5140, T5240, Server Architecture," http://www.sun.com/servers/coolthreads/t5140/wp.pdf, April 2008.

[15] Micron Technology Inc., "TN-41-01: Calculating Memory System Power for DDR3 Introduction," http://www.micron.com/support/part_info/powercalc.aspx, 2007.

[16] Micron Technology Inc. Various Datasheets: http://www.micron.com/products/dram/

References

[17] Rambus, "Challenges and Solutions for Future Main Memory," http://www.rambus.com/assets/documents/products/future_main_memory_whitepaper.pdf, May 2009.

[18] P. Chiang, M. Fung, "Dual-edge extended data out memory," US PATENT 5,950,223, September 1999.

[19] R. Barth, "2007 Test and Test Equipment," *2007 ITRS December Conference*, December 2007.

[20] H. Fujisawa, M. Nakamura, Y. Takai, Y. Koshikawa, T. Matano, S. Narui, N. Usuki, C. Dono, S. Miyatake, M. Morino, K. Arai, S. Kubouchi, I. Fujii, H. Yoko, T. Adachi, "1.8-V 800-Mb/s/pin DDR2 and 2.5-V 400-Mb/s/pin DDR1 Compatibly Designed 1Gb SDRAM With Dual Clock Input Latch Scheme and Hybrid Multi-Oxide Output Buffer," *IEEE International Solid-State Circuits Conference*, pp. 862-869, April 2005.

[21] C. Yoo, K. Kyung, G. Han, K. Lim, H. Lee, J. Chai, N. Heo, G. Byun, D. Lee, H. Choi, H.C. Choi, C. Kim, S. Cho, "A 1.8 V 700 Mb/s/pin 512 DDR-II SDRAM with on-die termination and off-chip calibration," *IEEE International Solid-State Circuits Conference*," Vol. 1, pp. 312-496, February 2003.

[22] C. Park, H. Chung, Y. Lee, J. Kim, J. Lee, M. Chae, D. Jung, S. Choi, S. Seo, T. Park, J. Shin, J. Cho, S. Lee, K. Kim, J. Lee, C. Kim. S. Cho, "A 512 Mbit, 1.6 Gbps/pin DDR3 SDRAM prototype with C_{IO} minimization and self-calibration techniques," *Symposium on VLSI Circuits*, pp. 370-373, June 2005.

[23] Y. Moon, Y. Cho, H. Lee, B. Jeong, S. Hyun, B. Kim, I. Jeong, S. Seo, J. Shin, S. Choi, H. Song, J. Choi, K. Kyung, Y. Jun, K. Kim, "1.2V 1.6Gb/s 56nm $6F^2$ 4Gb DDR3 SDRAM with hybrid-I/O sense amplifier and segmented sub-array architecture," *IEEE International Solid-State Circuits Conference*, pp. 128-129,129a, February 2009.

[24] F. Fishburn, B. Bush, J. Dale, D. Hwang, R. Lane, T. McDaniel, S. Southwick, R. Turi, H. Wang, L. Tran, "A 78nm $6F^2$ DRAM technology for multigigabit densities," *Symposium on VLSI Technology*, pp. 28-29, June 2004.

[25] C. Wintgens, "The 50-nm DRAM battle rages on: An overview of Micron's technology," http://www.eetimes.com, March 2009.

[26] H. Lee, D. Kim, B. Choi, G. Cho, S. Chung, W. Kim, M. Change, Y. Kim, J. Kim, T. Kim, H. Kim, H. Lee, H. Song, S. Park, J. Kim, S. Hong, S. Park, "Fully integrated and functioned 44nm DRAM technology for 1GB DRAM," *Symposium on VLSI Technology*, pp. 86-87.

[27] K. Kilbuck, "Main Memory Technology Direction," *Microsoft WinHEC 2007*, May 2007.

[28] B. Keeth, R.J. Baker, B. Johnson, F. Lin, *DRAM Circuit Design: Fundamental and High-Speed Topics, Second Edition*, Wiley-IEEE, 2008. ISBN 978-0-470-18475-2

[29] International Technology Roadmap for Semiconductor, *2007 Edition*, http://www.itrs.net/Links/2007ITRS/Home2007.htm, 2007.

[30] Samsung Semiconductor Inc. Various Datasheets: http://www.samsung.com/global/business/semiconductor/productList.do?fmly_id=690

[31] J. Handy, "Where Silicon is Headed and Why You Need to Know," Objective Analysis: http://www.media-tech.net/usa-09.html

[32] S. Kadivar, "New Memory Technologies: Evolving Toward Greener Solutions," Samsung Semiconductor Inc.: http://www.samsung.com/us/business/semiconductor/news/downloads/Green_Media_Event_SKadivar.pdf, March 2009.

References

[33] Hewlett-Packard, "Memory technology evolution: an overview of system memory technologies, technology brief, 8th edition,": http://h20000.www2.hp.com/bc/docs/support/SupportManual/c00256987/c00256987.pdf, April 2009.

[34] T. Jung, "Memory Technology and Solutions Roadmap," *Samsung ANALYST DAY*, 2005.

[35] R.J. Baker, *CMOS: Circuit Design, Layout, and Simulation, Revised Second Edition*, Wiley-IEEE, 2008. ISBN 978-0-470-22941-5

[36] L. Luo, J. Wilson, S. Mick, J. Xu, L. Zhang, P. Franzon, "3 gb/s AC coupled chip-to-chip communication using a low swing pulse receiver," *IEEE Journal of Solid-State Circuits*, vol. 41, Issue:1, pp. 287-296, January 2006.