A 22nm 20Mb Embedded MRAM with 5Gbps Read and 1Gbps Programming

Presenter: Nick Hendrickson

Jiancheng Huang, Doug Smith, Chun-Tai Cheng, Nilesh Gharia, Jy-Hong Lin, Charles Farmer, Wen-Chun You, Radu Avramescu, Tien-Wei Chiang, Michael Jaggers, Kuei-Hung Shen, Jack Guedj, William J. Gallagher, Harry Chuang

numem

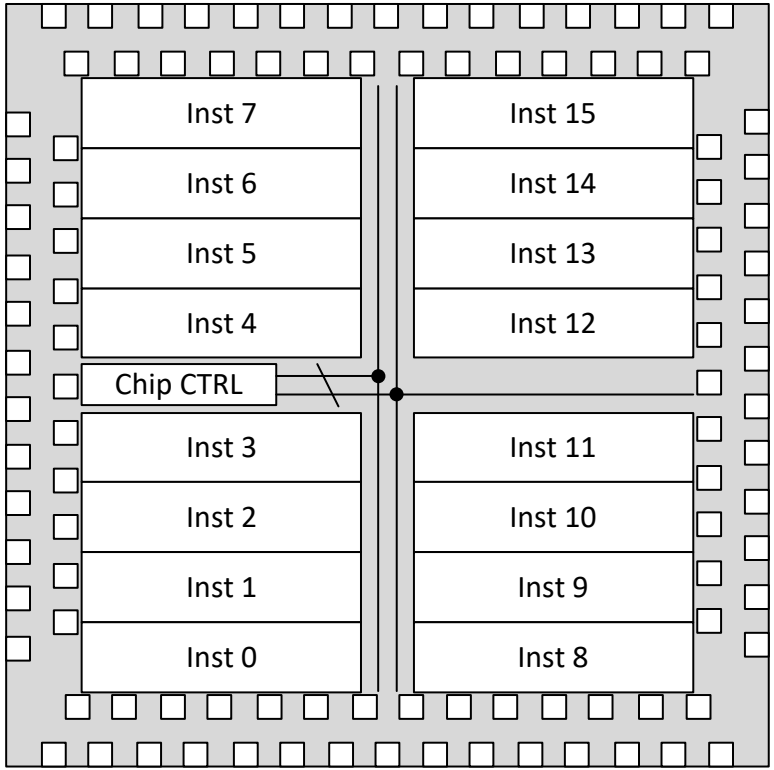
Introduction

- The data presented here is for Numem's first generation TSMC 22nm MRAM
- The primary goal of this architecture is for high performance, high reliability embedded MRAM IP
- The testchip along with our custom test platform provides a low cost, scalable means to test large amounts of memory in parallel while still allowing discreet analog access to every cell on the chip.

Numem Confidential and Proprietary Information..

Chip Overview

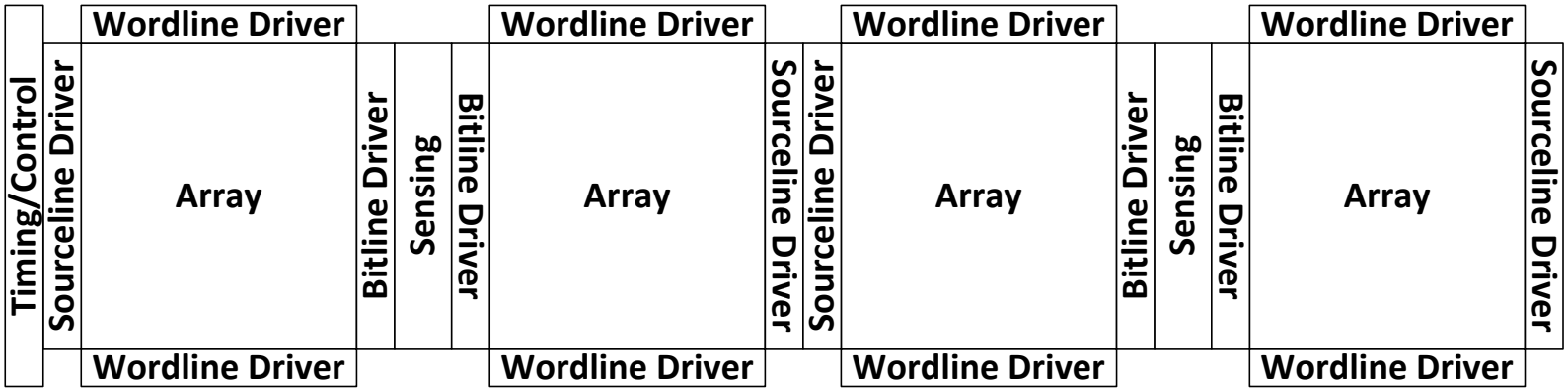
- The testchip makes use of 16 high-performance, high-density MRAM instances
  - TSMC 22nm
  - 0.0456um^2 bitcell
- Numem's standard testchip interface enables very high throughput, flexibility and control at modest pincounts
  - 32 pin digital interface
  - 3 analog pins
  - Remainder are various power/ground
  - 100 pin total



Numem Confidential and Proprietary Information..

Instance Overview

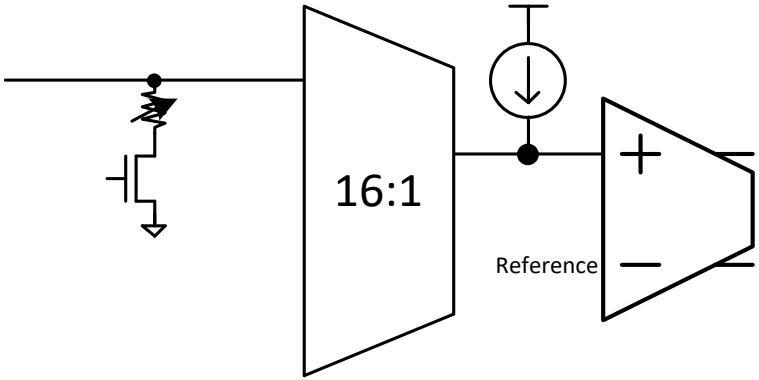
- Daisy chainable block; 2 deep for demonstration purposes
  - Center mounted Sense Amplifiers to share between two MRAM arrays without additional metal expense
  - 512WL x 640BL array
  - 40b Data Word
  - Easily adaptable to 80b Data Word for higher throughput and area efficiency
- Each instance interfaces by means of a standard SP SRAM interface
  - Simplifies integration into existing systems
  - Provides sufficient bandwidth to not bottleneck operation
  - Incurs no meaningful latency penalty at system level



Numem Confidential and Proprietary Information..

Read Architecture

- Reference generation is the most critical element of an MRAM read architecture. There reference must track:
  - PVT
  - Bitline resistance/position
  - Sourceline resistance/position
  - Wordline voltage / access device resistance
  - Without compensation of these terms, a state separation of just 2.4 sigma is left from a starting separation of 8 sigma!
- **Numem's patented reference generation methodology has demonstrated near perfect compensation for all of these variations, retaining nearly all of the theoretical read window**
- A forced current sense approach is able to settle within 6ns
  - Allowing for address/data propagation and sense resolve time, the total access time is 8ns, or 5Gbps



	Sigma
Median State Seperation*	7.9
Uncompensated Bitline resistance	1.6
Uncompensated Sourceline resistance	2.1
Uncompensated Wordline Voltage	1.8
Uncompensated seperation	**2.4**

* (medianRH - medianRL) / (stdRH + stdRL)

Numem Confidential and Proprietary Information..

Programming Performance

- All 40 bits are programmed in parallel
- Verify is implemented to eliminate program soft errors
- Single pulse SER of <100PPM is achieved @ 32ns
  - Just 10% overdrive voltage required during programming
  - Increasing program time to 64ns only reduces the VPRG by 6%
  - Lower programming time and increased programming voltage generally improves overall power usage

Program Voltage and Time vs Soft Fail Rate (PPM)

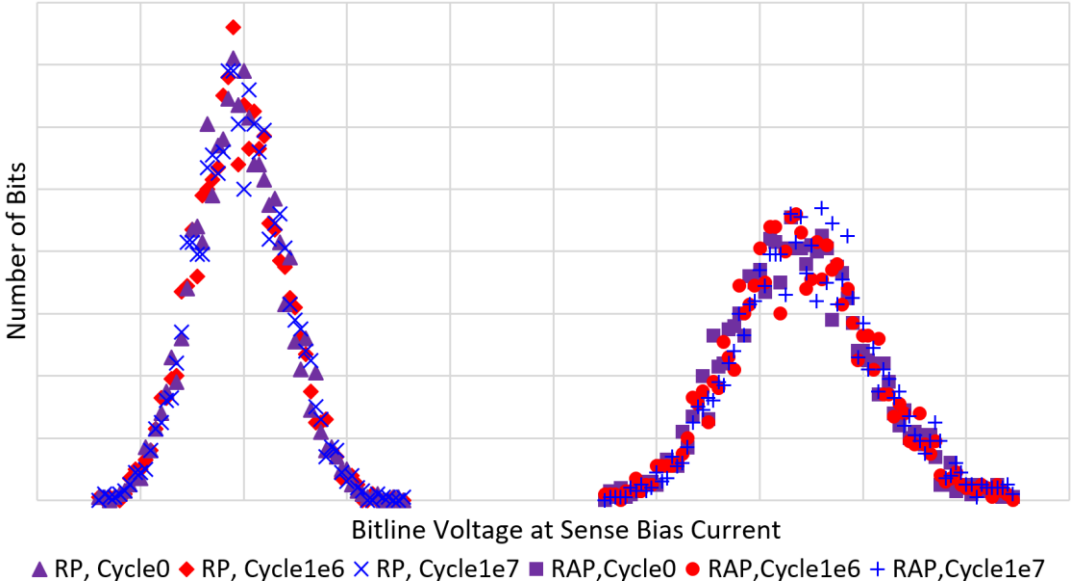
VPRG	32ns	64ns	96ns	128ns
1.14x	7	0	0	0
1.12x	18	0	0	0
1.10x	45	0	0	0
1.08x	113	2	0	0
1.06x	289	6	1	0
1.04x	735	29	4	1
1.02x	2097	174	36	11
1.00x	6682	1193	408	184

* VPRG scaled to nominal process maximum voltage

Numem Confidential and Proprietary Information..

Aging Results

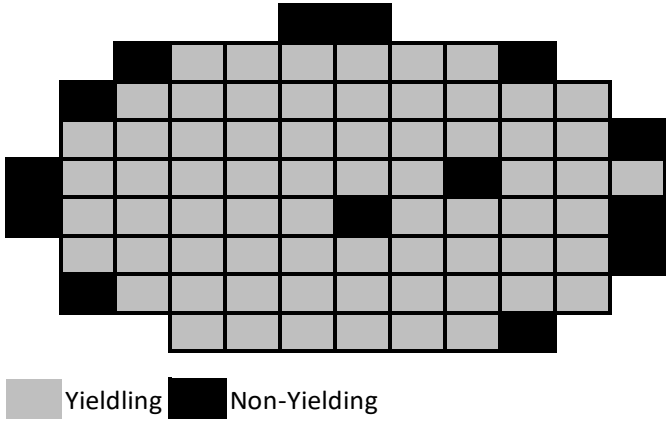
- The bitline voltage data shown quantifies the total resistance of the cell including RMTJ, RACC, RBL, RSL, etc.
  - Chip level features are mostly compensated for in the test itself
  - While actual sense window distributions are better compensated, this provides a detailed analog view of cell aging
- Across cycling, virtually no drift in resistance for either Rp or Rap is observed
  - These values have been measured at time 0, 1e6, and 1e7 with no noticeable change in median or sigma values
  - This is important in order to ensure reference placement / read windows are constant as the device ages



Numem Confidential and Proprietary Information..

Yield

- Local yield on healthy die are very good, with the process moving into production fabs
- Column and Wordline repairs are optional, but generally recommended to optimize yield
  - On this testchip, over 90% of die show no need for either repair mechanism anywhere on the die
- Bit level repairs are reasonable and being actively monitored as the process continues to mature
  - This design implemented with 16 bit repairs per Mb; nearly all otherwise healthy die yield within that repair
  - Spatial mapping shows failing bits to be randomly distributed throughout the memory
  - Bit defectivity includes: opens, shorts, out-of-distribution resistance
  - All of these defect types are repaired using Word repair mechanism



Numem Confidential and Proprietary Information..

Summary

- This test chip demonstrates a current generation MRAM technology, production ready, delivering on some of the most aggressive performance predictions
- A simple SP SRAM interface is ideal to harness the bandwidth and latencies available for embedded implementations
- High performance and low power for both reads and writes positions this MRAM as both a superior NVM replacement as well as a real SRAM competitor for low power, low speed sockets
- Yields are already sufficient for volume manufacturing and continue to improve

Numem Confidential and Proprietary Information..