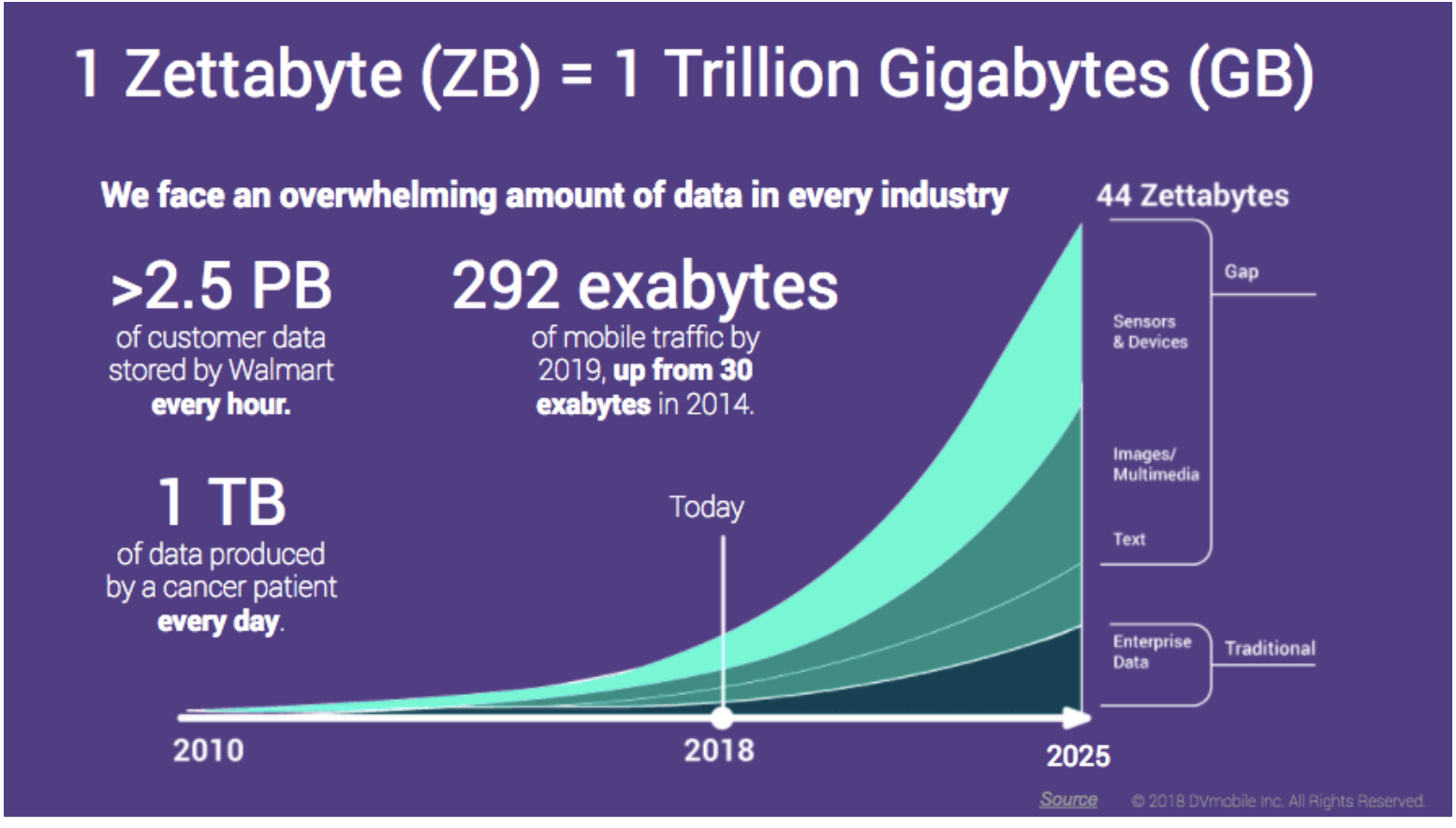
When it comes to Emerging Memories ……One Size Doesn't Fit All!!

August 8, 2019

numem

Overwhelming Deluge of Data



…Driving the Need for more efficient Memory and more distributed processing

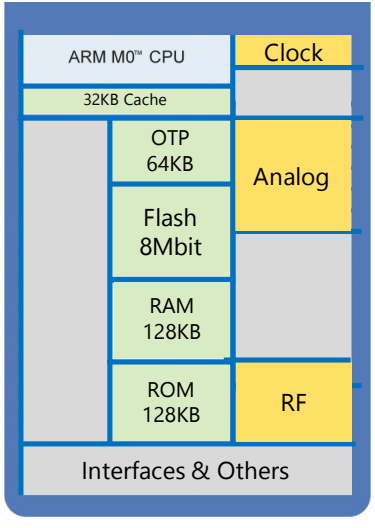
numem

Edge Processing: Increased Edge Compute/Memory

- Due to the increased data collection, it is no longer practical and power efficient to move all the data to the Cloud
- Higher Processing at the Edge
- Higher Non-Volatile Memory to store
  - Program, Models/Coefficients, increased amount of Data collected
- Higher Volatile Memory for AI/Signal Processing
- Ultra-low Power to extend Battery Life and to reduce thermal issues to reduce cost and form factor

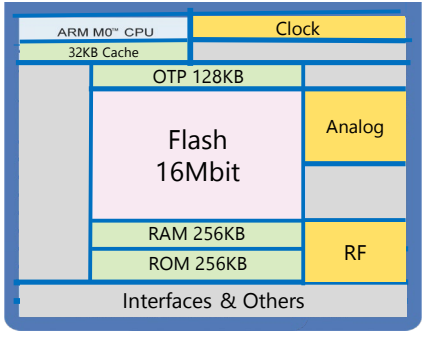
➔ **More than ever, it is critical to have efficient memory in edge nodes**

Resistive RAMs Provide a Simpler/Smaller Memory



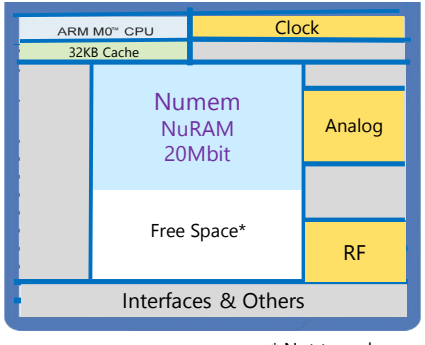
IoT SoC

Next Gen. Option-I --- > Current Memory Tech.



* Not to scale

Next Gen. Option-II --- >Numem NuRAM



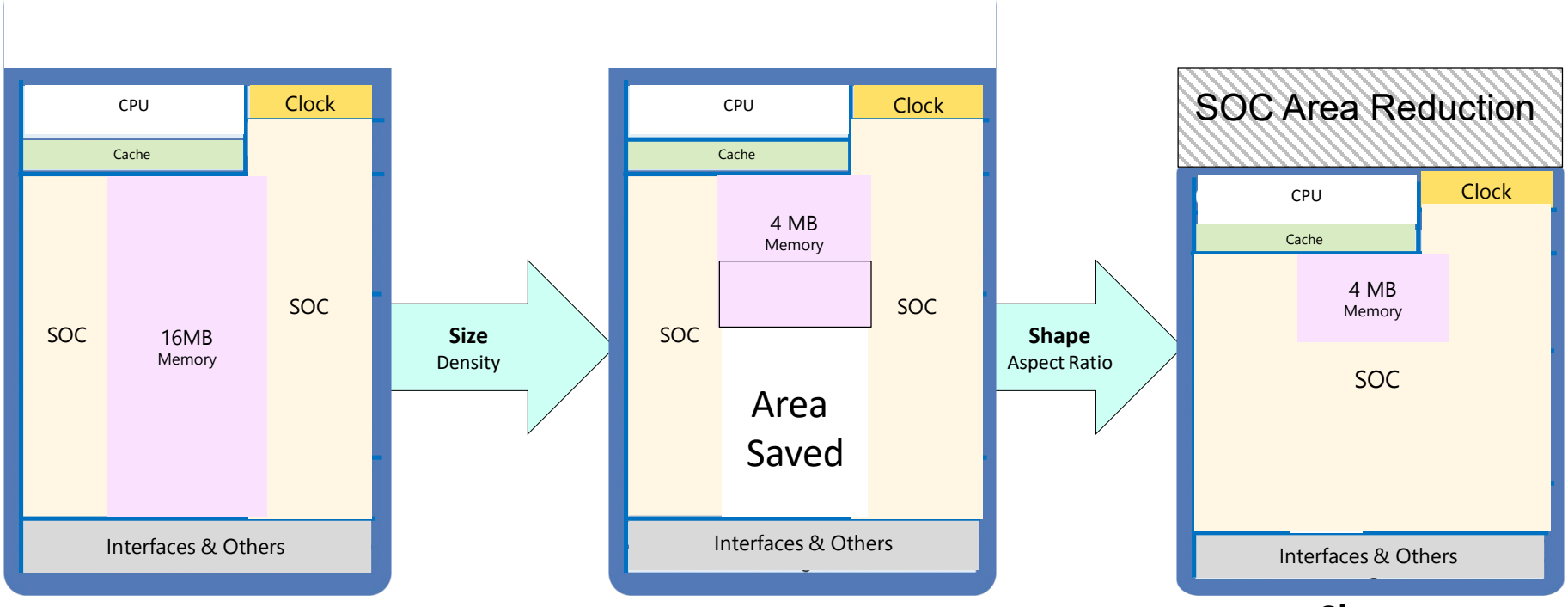
* Not to scale

Resistive RAMs such as MRAM, RRAM, PCRAM, enable Simplified & Smaller Unified Memory

Optimization: Significant Improvements in Speed/Power/Area

- Architecture
  - Memory Architecture: Read/Write Circuits, Decoders, Array
  - SOC Architecture and Distribution of Memory throughout the SOC
  - Memory Subsystem: BIST, Failure Correction (ECC, Repair, Calibration)
- Optimized Circuit Design affecting even more RRAM/PCRAM/MRAM (newer technologies)
- Design Customization including:
  - Size
  - Shape/Aspect Ratio (Column Muxing)
  - Array Subdivision/ Bit line length
  - Pipelining
  - Reflow/Non-Reflow Bitcells

Area: Exact Sizing and Shape Optimization



Size

4Mb required vs 16 Mb
Square shape

Shape

Rectangular shape required
due to shape/routing congestions

Not to Scale

Area: Exact Sizing and Shape Optimization

- Standard IP Cores are usually One-size-fits-all: e.g.16Mb & 32Mb
- Example: you need a 4Mb & only 512K Reflow Capable (~20% larger)
- Tally up delta size 4x1.2-0.5x(1-1.2)..... **4.7x larger**
- Shape/Routing Congestions also result in larger SOC: **~30% larger**
  - It is not always about the memory best area but about best SOC area
- Memory Design Architecture also affects size enormously: **>2x delta**

➔Tally up: **overall size delta could be in excess of ~13.4x** in this example (4.7/0.7*2)

Speed/Power Optimization

- The Memory design can significantly affect Speed/Power
  - Array Subdivision (Bitline length)
  - Pipelining
  - Optimized Circuit Design (Sense Amp, Decoder, Data Pipe, etc..)
- Shorter Bitline Integration time can improve performance by **>2x**
  - 5-30% larger area depending on optimization
- Design architecture is of paramount importance for best possible performance especially in newer technologies like RRAM/MRAM
  - Read Access times vary can widely from 2.5ns to 20+ns typ so by **8x**
- Pipelining achieves significantly higher bit rate (if acceptable), **2x** per stage

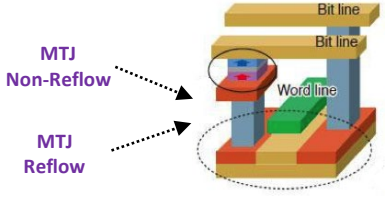
➔All and all optimization can reasonably **improve performance by >16x**

Optimization/Customization Improvements Summary

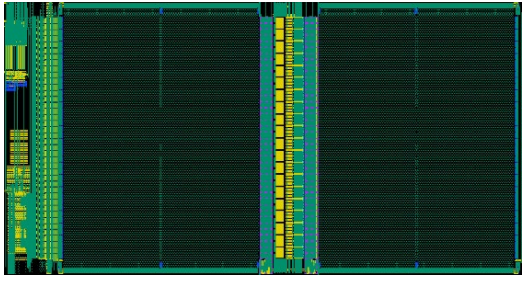
- >13x Smaller Area
- >16x better Speed/Power
- 50X Faster Time to Market through Design Automation SW
- More Reliable Product through Validation/Characterization SW

Numem: Best-in-Class Memory Subsystem IP Cores

Patented Architecture and Design (Smaller size, higher Perf./Lower Power)

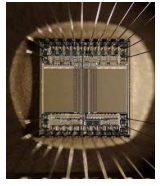


MRAM/RRAM/PCRAM
Foundry Technology Bitcell

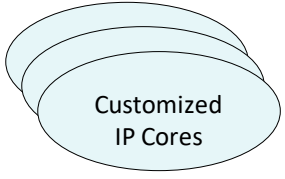


Test & Qual

Test Chips



Design Automation SW



- **Optimized Standard & Custom Memory Subsystem IP Cores**
  - Validated on MRAM through tape-out of 19 devices to date and extensive testing
  - Our memory are non-volatile and 2-3x smaller than SRAM
- **Design Automation SW for Memory Customization & Validation/Characterization**
  - Yields much greater PPA improvements and higher reliability
- **Memory Chiplets** (die) for use in MCP SOCs

Thank You