HETEROGENEOUS INTEGRATION ROADMAP

2019 Edition

Chapter 17: Test Technology

Section 05: Memory Test

http://eps.ieee.org/hir

The HIR is devised and intended for technology assessment only and is without regard to any commercial considerations pertaining to individual products or equipment.

We acknowledge with gratitude the use of material and figures in this Roadmap that are excerpted from original sources.
Figures & tables should be re-used only with the permission of the original source.

IEEE ELECTRONICS PACKAGING SOCIETY · ELECTRON DEVICES SOCIETY · IEEE Photonics Society · ASME SETTING THE STANDARD · semi

Section 5: Memory Test

There have been significant changes in the Memory environment in recent years. Up until 2003, DRAM bits comprised approximately 90% of bits shipped per month. By 2009, NAND had become the dominant form of memory and comprised 85% of monthly bits shipped. During the same period, NOR Flash has largely migrated from parallel interface devices to serial interface devices with extremely small form factors in order to reduce PCB size, complexity and power. With the introduction of mobile smart phones in 2007 and tablets in 2010, the traditional dominance of PC DRAM has been eroded by lower-power LPDRAM which is required for longer battery life. New generation memory types such as PRAM, RRAM, STT RAM, and CBRAM along with 3D multilayer instantiations of NAND will further change the memory environment over the next decade. The continuing shift to a battery-powered wireless environment will continue to drive changes in memory usage. Existing memory types will likely not be replaced, but will be used in joint solutions with newer memory types.

Memory density has kept pace with Moore's Law since the first DRAM was manufactured in 1969, but lithography cycles as well as the performance roadmap are expected to push out for litho nodes less than 20nm, so the typical 2-year technology pace is forecasted to stretch to initially 3 years, and then 5 years later in the roadmap. Multi-layer 3D NAND technologies allowed the use of longer gate lengths, so the 2-year technology pace should continue for the near term.

From a test perspective, most memory will be structurally tested at wafer test utilizing low-pin-count interfaces of 4 to 10 pins per device. Structural test, when used, will likely include a self-test performance validation.

Table 1: Memory Test Requirements

	2018	2019	2020	2021	2026	2031
DRAM Characteristics						
Capacity (bits) [1]	16G	16G	32G	32G	64G	128G
I/O data rate (Gb/s)						
PC DDRx	3.2	4.4	4.4	6.4	8.4	8.4
GDDRx	10.0	12.0	16.0	16.0	16.0	16.0
LPDDRx	4.0	4.8	4.8	4.8	8.4	8.4
Hybrid Memory Cube (HMC [SerDes])	15.0	15.0	15.0	15.0	25.0	30.0
Wide IO	0.3	1.1	1.1	1.1	2.2	2.2
High Bandwidth Memory (HBM)	2.0	2.6	2.6	2.6	3.0	3.6
NAND Characteristics						
Capacity (bits) [2][3][4]	1T	1T	2T	2T	16T	64T
Maximum I/O data rate (Gb/s)	0.53	0.53	0.53	0.67	1.07	1.60

Manufacturable solutions exist, and are being optimized

Manufacturable solutions are known

Interim solutions are known

Manufacturable solutions are NOT known

Notes:

1. DRAM bit capacity per die
2. 3 bits per cell introduced in 2009
3. 4 bits per cell introduced in 2012
4. 3D multi-layer introduced in 2014

DRAM

Historical trends show that PC DRAM has doubled its performance every 5 years and will reach a data rate of 8.4 Gb/s per I/O in 2022 with DDR6. However, DDR4 appears to be the end of the DDRx era as there are no further enhancements of the DDRx architecture beyond DDR4 in definition or development. DDR4 itself has severe PCB design restrictions in order to meet the I/O performance requirements, so an enhancement of the current DDRx single-ended interface will present additional challenges and constraints. The Wide IO memory architecture is targeted to mobile applications such as phones and tablets and is an evolution of the DDRx that decreases I/O bit rates while expanding the number of I/Os up to 512. Hybrid Memory Cube (HMC) is targeted to servers, where it provides very high performance over a SERDES interface, though at increased cost. High Bandwidth Memory (HBM) is similarly

targeted to graphics video cards and applications. Based on existing roadmaps, DDR3/4 will continue to serve the PC market for the foreseeable future. LPDDRx and GDDRx DRAM families will continue to be a driver for the near future.

DRAM will become increasingly difficult to scale in sub-20nm nodes, and transistor wear-out will increase the frequency of errors. On-chip error correction and memory management will likely become a requirement before 2020. Dynamic failure detection, analysis, and repair will become necessary over the product life. To enhance test productivity, new test-oriented architectures will be required. On-chip correction may also change the DRAM test paradigm.

Maintaining high ATE test parallelism is required over the roadmap period to manage test cost. However, probe card performance test at high parallelism may be a challenge at high GT/s due to the interface routing complexity required. These challenges will ultimately drive the need for die self-test.

Flash

NAND will double in density ever year in the short term and slow to a doubling every 2 years. The doubling every 2 years will be faster than the projected lithography node migration due to the increase in the number of bits stored in a single memory cell from one and two to four on some cell types. Use of error correction allows greater uncorrected error rates and enables increased bits per memory cell. NAND bus width has continued to be dominantly 8-bit with a decreasing number of products at 16-bit I/O. As large amounts of NAND are being consumed in Solid State Drives (SSD), a new NAND interface may emerge that is more optimized for SSD use.

The need for internal voltages that are 3-8 times the external supply requirements is expected to continue in the test process, driven by the hot-electron and Fowler-Nordheim charge transport mechanisms. Increased absolute accuracy of supply voltages will be required in the future due to the trend toward lower voltages, but percentage accuracy relative to the supply voltage will remain a constant. I/O voltage decreases are pushing the operational limits of standard tester load circuits; new methods will be required in the future.

Wafer test generally does not require the performance of package test, but error detection, error analysis, and redundancy processing are required.

NOR memory density is expected to increase slowly over the roadmap period and remain flat toward the end of the roadmap. NOR has been transitioning from a parallel to a serial interface since 2007 to reduce package size and power. Further increases in NOR performance along with increasing test requirements are not expected.

Stacking of various types of Flash and other memory and/or logic components in a single package has become standard and is expected to continue. Multiple die within a package has complicated the package test requirements and has increased pin count and number of DUT power supplies required. Data and clock rates for flash will increase, but there is expected to be a wide variability in the requirements based upon the end application.

Embedded Memory

Embedded memory consumes greater than 80% of transistors in many MPU and SoC designs and will scale with the increase of transistors in these devices. Embedded Flash and DRAM bits will not match the density of standard DRAM and NAND. Newer memory types such as RRAM or STT RAM may become embedded over the course of the roadmap.

To enhance test productivity, new test-oriented architectures and/or interfaces will be required. Built-in self-test (BIST) and built-in self-repair (BISR) will be essential to test embedded DRAM and Flash memories cost effectively. The primary test algorithms for Flash memories will continue to be Read-disturb, Program-disturb, and Erase-disturb while March tests with all data backgrounds will be essential for embedded DRAM.

Considerable parallelism in test will be required to maintain test throughput in the face of rising memory densities. In some cases, test is made cost-effective by double insertion of devices rather than testing both logic and embedded memories on the same tester. In double insertion, embedded Flash and DRAM are tested and repaired on a memory tester, while the logic blocks are tested on a logic tester.