Microelectronic Engineering 164 (2016) 75–87

Contents lists available at ScienceDirect

Microelectronic Engineering

journal homepage: www.elsevier.com/locate/mee

Uprising nano memories: Latest advances in monolithic three dimensional (3D) integrated Flash memories

Mojtaba Joodaki

[a] Department of Electrical Engineering, Ferdowsi University of Mashhad, 9177948974 Mashhad, Iran
[b] Sun-air Research Institute, Ferdowsi University of Mashhad, 9177948974 Mashhad, Iran

ARTICLE INFO

Article history:
Received 5 November 2015
Received in revised form 3 July 2016
Accepted 26 July 2016
Available online 27 July 2016

Keywords:
NAND Flash memory
3D integrated Flash memory
Bit-cost scalable (BiCS) Flash memory
Vertical recess array transistor (VRAT) Flash memory
Vertical gate NAND (VG-NAND) Flash memory

ABSTRACT

Flash memory industry has showed remarkable steady progress during the last few years. This achievement is owed to the development of the 3D NAND Flash structures. This paper reviews the latest advances in the monolithic 3D integration of the NAND Flash memory. It highlights key technical challenges and uses important aspects and characteristics of various designs in order to illustrate mechanisms that overcome the technical barriers. Furthermore, the most promising solutions are marked by comparing advantages and disadvantages of different designs. Finally, future prospects and expected market demand of NAND Flash memory are addressed.

Contents

1. Introduction

A major part of semiconductor industry is devoted to memory products. In 2013, the worldwide semiconductor sales reached $305.6 billion, an increase of 4.8% from those of 2012. Although memory was the second semiconductor category by sales reaching $67.0 billion in 2013, it was the fastest growing segment, increasing 17.6% in the same year [1]. This growth is also evident in 2014, when the worldwide semiconductor sales reached $335.8 billion, 9.9% higher than those of the previous year [2]. That was a record year for semiconductor revenues. Again, memory was the highest growing segment with 18.2% increase in 2014 (total sales of $79.2 billion) [2]. Last year the situation was very similar to that of 2014 [3,4]. The semiconductor global sales reached $335.2 billion from which $77.2 billion belong to the memory chips [4]. In spite of the limited growth in 2015, a slightly positive market growth in 2016 and a moderate one in 2017 are predicted [5].

E-mail address: joodaki@um.ac.ir.

http://dx.doi.org/10.1016/j.mee.2016.07.009
0167-9317/© 2016 Published by Elsevier B.V.

The industry owes this impressive growth to Moore's Law, which has dramatically reduced the cost of memory devices over the past few decades. It has also produced products of ever-higher capacity and eventually has provided the consumers with enormous amounts of memory at a much lower cost. For instance, the cost of Flash memory per megabyte (MB) has fallen by a factor of 60,000 in 20 years (from 1987 to 2007). This corresponds to halving the cost by doubling the density every 15 months [6]. Interestingly, the Flash memory cost reduction trend has kept on going and is expected to continue, at least, for the few years coming. The cost of Flash memory in 2007 ($12 per GB) has already fallen by a factor of 30 in today's 1 TB devices and is expected to reduce by 50% (30 cents per GB) this year, when each solid-state memory (SSM) chip will offer up to 2 TB of storage [7].

Semiconductor memories are often categorized according to their volatility, alterability, readout destructivity, access scheme (e.g. random or block-wise), and refreshing of contents. Fig. 1 briefs the different memory types that are presently in frequent use. Volatile memories need power supply to preserve the stored data. DRAM and SRAM are the two dominant types of volatile memories in the market. The term RAM (random access memory) means a memory device in which cells are accessed at random for read, erase and programing operations. A SRAM is mainly used for two purposes: as cache memory to facilitate an interface with the CPU at speeds not achieved by DRAM (memory wall problem) and to take the place of DRAM in systems that consume very low power. In a DRAM cell, each bit of data is stored on a separate capacitor. Since the capacitors will be slowly discharged by time and this may cause information loss, they must be refreshed periodically.

Nonvolatile memories, by contrast, maintain the stored data even when the power supply is off. Mostly a nonvolatile memory should have an unpowered retention time bigger than 10 years; however, this parameter can be reduced depending on the specific memory technology used and its application. The nonvolatile property empowers the memory system to cover a wide range of applications, from consumer and automotive to computer and communication products. Nonvolatile memories are often categorized based on the degree of functional flexibility in altering the stored data. For instance, a ROM is a type of semiconductor memory in which the information is written only once and then it cannot be changed. ROM programming is done during the manufacturing process, whereas in a PROM, the user can program the device just before its insertion into a system, though it is not possible to alter the data. The customer can program these memories by using a special apparatus called PROM programmer. A PROM consists of an array of fuse links some of which are "burnt" during the programming operation to have a specific data pattern. In EPROM, the limitations of ROM and PROM have been reduced, as it can be electrically written and then erased later by exposing the device to UV light. This demands the insertion of a quartz window on the top of the package and displacement of the memory chip from the circuit board to insert it into a special UV eraser. In an EEPROM, by contrast, both program and erase operations are performed using an electrical voltage. E^2PROM is almost the same as EEPROM; however, in E^2PROM, information is altered one byte at a time, whereas in EEPROM information is erased in blocks.

These examples reveal that in order to boost memory sales revenue and remain competitive, alternating post-Flash devices with various storage mechanisms and reliability issues are necessary. Among a relatively large number of candidates magnetic RAM (MRAM), ferroelectric RAM (FeRAM), phase change RAM (PCRAM), and resistive RAM (RRAM) are either nearer to production or already in production for some applications.

Flash memory is an EEPROM that can be erased in large blocks or sectors and then can be reprogrammed in either byte/word or pages. There are several variants of Flash memory that can be classified according to access scheme, their being parallel or serial, or in terms of programming and erasing mechanisms employed by them, such as channel hot electron (CHE), hot holes (HH), Fowler-Nordheim tunneling (FN), and source-side hot electron (SSHE). However, two types of them dominate the current Flash memory market: NAND Flash (suitable for mass data storage applications), and common ground NOR

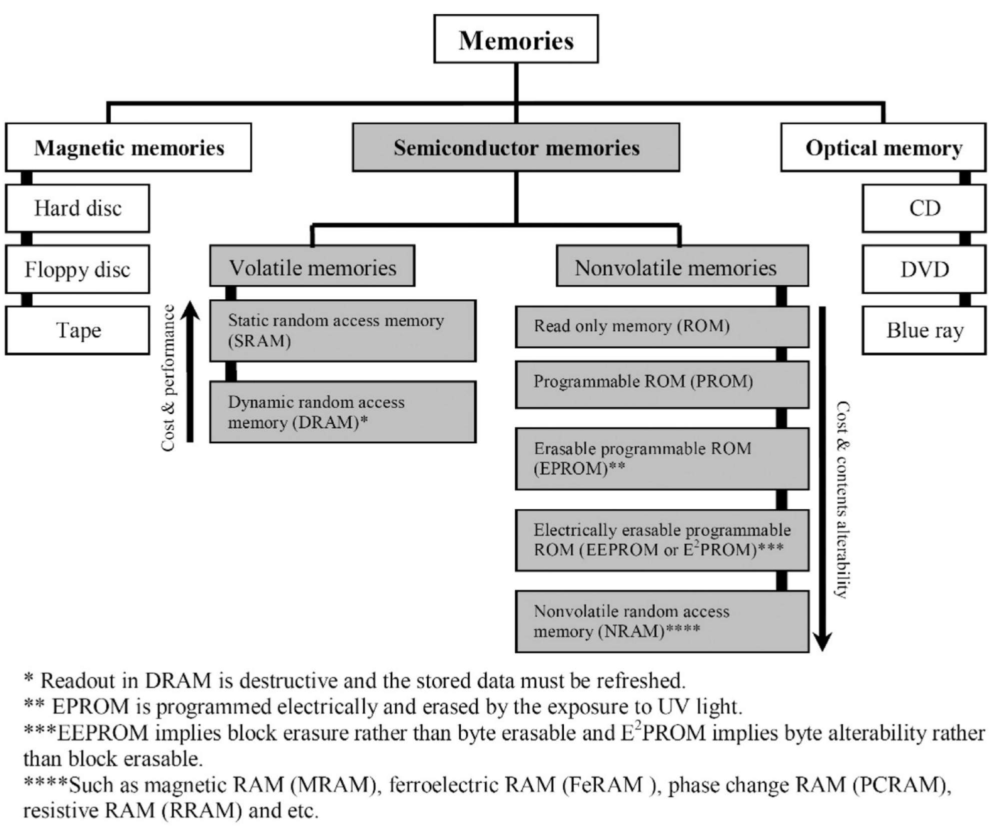

Fig. 1. Semiconductor memory categories currently in frequent use.