FLASH MEMORY TECHNOLOGY EVOLUTION

Roberto Bez and Emilio Camerlenghi

STMicroelectronics, Central R&D
Via C.Olivetti 2, 20041 Agrate Brianza (Milan), Italy

The most relevant phenomenon of this last decade in the field of semiconductor memories has been the explosive growth of the Flash memory market, driven by cellular phones and other types of electronic portable equipments (palm top, mobile PC, mp3 audio player, digital camera and so on). Moreover, in the coming years portable systems will ask even more non volatile memories either with high density and very high writing throughput for data storage application, or with fast random access for code execution in place. The strong consolidated know-how (more than ten years of experience), the flexibility and the cost make the Flash Memory a largely utilized, well-consolidated and mature technology for most of the non-volatile memory application. Today Flash sales represent a considerable amount of the overall semiconductor market.

Although in the past different type of Flash cells and architectures have been proposed, today two of them can be considered as industry standard: the common ground NOR Flash, that due to its versatility is addressing both the code and data storage segments, and the NAND Flash, optimized for the data storage market.

In this paper the development of the NOR Flash memory technology will be presented with the aim of describing both the basic functionality of the memory cell used so far and the main cell architecture today consolidated. The NOR cell is basically a floating-gate MOS transistor, programmed by channel hot electron and erased by Fowler-Nordheim tunnel. The main reliability issues, like charge retention and endurance, will be discussed, together with the understanding of the basic physical mechanisms responsible for. Most of these considerations are valid also for the NAND cell since it is based on the same concept of floating gates MOS transistor.

Furthermore an insight into the multilevel approach, where two bits are stored in the same cell will be presented. In fact the exploitation of the multilevel approach at each technology node allow the increase of the memory efficiency, about doubling the density at the same chip size, enlarging the application range and reducing the cost per bit.

Finally the NOR Flash cell scaling issues will be covered, pointing out the main challenges. The Flash cell scaling has been demonstrated to be really possible and to be able to follow the Moore's law down to the 130 nm technology generations. The technology development and the consolidated know-how is expected to sustain the scaling trend down to the 90 nm and 65 nm technology nodes as forecasted by the ITRS roadmap. One of the crucial issues to be solved to allow cell scaling below the 65 nm node is the tunnel oxide thickness reduction, as tunnel thinning is limited by intrinsic and extrinsic mechanisms.