NVM future trends – from floating gate to trapping devices

B. Eitan, I. Bloom, G. Cohen, M. Janai, E. Lusky, and A. Shappir
Saifun Semiconductors Ltd., Israel

Non-volatile memory (NVM) has experienced tremendous growth in the past five years. This trend is expected to continue in the coming years. The workhorse of most NVM products is the floating gate (FG) device. The last few years have focused on scaling and increasing the number of bits in a multi-level cell (MLC) and on new emerging technologies, either based on trapping-concepts (NROM, nano-crystals and SONOS) or based on new physics and material concepts (FERAM, MRAM and phase-change).

The scaling of area per bit in the FG NVM memories is focused on traditional shrinking of the lithography and increasing the number of bits to two (MLC). The lithography scaling path suffers from the non-scalability of the voltages, as the tunneling oxide has only changed from 10 nm in 1980 to 8.5 nm in 2005. Another great limitation is the gate to FG coupling ratio: scaling the cell reduces the FG capacitance and hence results in a smaller number of stored electrons, limiting the reliability of the MLC devices. These scaling difficulties in the FG devices are creating a significant opportunity for the emerging NVM technologies.

Among the trapping devices, mainly the NROM family, scaling is progressing faster than the lithography road map. This paradigm shift is enabled by increasing the number of bits per cell from two to four and reducing the cell size below the $4F^2$ lithography limit. In this presentation, we will show that this faster dowscaling is enabled by the NROM device physics, including:

- program and erase with hot carriers that reduces the operating voltages ;
- tunnel oxide scaling (to 3 nm) based on storage on a dielectric rather than on a conductor;
- coupling ratio of one and no FG-related cross talk;
- and a very simple manufacturing process with great similarity to CMOS scaling flows.

The NROM concept, based on trapping in the ONO dielectric, is already in volume production for code, data and embedded applications. Of the same family, the nanocrystal memories are having major process integration problems and offer no advantage over the ONO-based devices. The SONOS-based devices rely on tunneling program and erase, hence suffer from the very high voltage requirements.

The second category of technologies is based on new physics and material concepts. For this category to make an impact they must have a small cell size with at least two bits per cell. Unfortunately all these new concepts have more than one element per bit. They also suffer from material science complications. This makes them unlikely candidates to replace any of the technologies that are already in production.

The system architecture that supports the memory is becoming an important part of the scaling path. Concepts like error detection and correction, ware leveling and bad block management will become a must in the future, when redundancy techniques will no longer be sufficient. We will discuss how the choice of trapping technology makes a major impact on the simplicity and efficiency of the system integration.

The scaling limitation of the FG NVM products is already very visible. The NROM, as a charge trapping based technology, avoids these limitations. Together with the proven ability to deliver four bits per cell and a good system compatibility, NROM is an interesting example for the CMOS technologies that "straight forward" scaling is not the only option.