

Semiconductor Technology from A to Z

Digital technology

The Flash Memory Cell	3
<i>Structure of the Floating-Gate Cell</i>	3
<i>Programming and Erasing</i>	3
<i>NOR vs. NAND Architecture</i>	4
<i>Multi-Level Cell (MLC/TLC)</i>	4
<i>Comparison: DRAM, SRAM, and Flash</i>	4

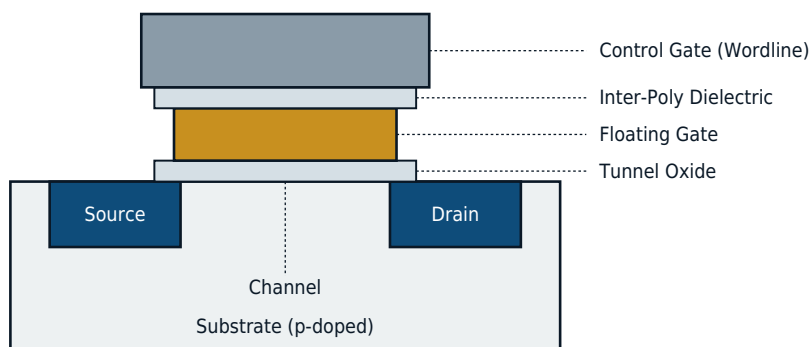
The Flash Memory Cell

Structure of the Floating-Gate Cell

A flash memory cell is essentially a MOSFET with an additional, electrically isolated gate electrode – the floating gate. It sits between the regular control gate and the channel, fully enclosed by silicon dioxide: separated from the channel below by the thin tunnel oxide (a few nanometers), and isolated from the control gate above by a thicker interpoly oxide.

Because the floating gate has no external electrical contact, charge once placed on it remains stored for years – this is the basic principle of non-volatile storage.

The Floating-Gate Cell
Structure in cross-section



The floating gate is completely surrounded by oxide and therefore electrically isolated.
The stored charge remains in place for years.

Programming and Erasing

To program a cell (placing charge on the floating gate), either hot-electron injection or Fowler-Nordheim tunneling is used, depending on the cell type: with hot-electron injection, a high drain-source field accelerates electrons strongly enough that some gain sufficient energy to overcome the oxide barrier and become trapped on the floating gate.

Fowler-Nordheim tunneling instead uses a very high electric field across the tunnel oxide, allowing electrons to tunnel through the barrier quantum-mechanically – this mechanism is typically used for erasing, where charge is removed from the floating gate again. The stored charge measurably shifts the cell's threshold voltage, allowing "0" and "1" to be distinguished during readout.

NOR vs. NAND Architecture

The way individual cells are wired into an array distinguishes two basic architectures. In NOR flash, each cell is individually connected in parallel to the bitline and wordline, similar to a NOR gate – this enables fast, random access to individual bytes, but costs area.

In NAND flash, several cells (typically 32–128) are connected in series between the bitline and source line, as in a NAND gate; this saves considerable area and enables today's high storage densities, but only allows page-wise rather than random access. This is why NAND flash dominates in USB drives, SSDs, and memory cards, while NOR flash is used where fast direct access to individual addresses is required, such as in firmware storage.

Multi-Level Cell (MLC/TLC)

Instead of distinguishing only between two charge states (0/1), the amount of charge on the floating gate can be graded more finely. An SLC cell (single-level cell) stores 1 bit, an MLC cell stores 2 bits across four distinguishable charge levels, and a TLC cell stores 3 bits across eight levels.

This multiplies the storage density per cell but reduces the margin between levels – the cells become more sensitive to charge loss and require more elaborate error correction, and the number of possible write/erase cycles also decreases.

Comparison: DRAM, SRAM, and Flash

The three memory types differ fundamentally in volatility and intended use. DRAM stores charge in a capacitor and must be refreshed continuously, losing its content immediately on power loss – but it is fast and serves as [main memory](#).

SRAM stores its state in a bistable circuit of six transistors, is even faster but more area-intensive, and is used for example as [cache memory](#). Flash, by contrast, retains its content without power but is significantly slower to write than DRAM and SRAM, making it suited for persistent data storage rather than main memory.