**Flash Memory**

Table of Contents

[Introduction 2](#_Toc65536665)

[Floating-Gate Transistors 4](#_Toc65536666)

## Introduction

Flash memory is an electronic, non-volatile computer storage medium that can be electrically erased and reprogrammed. It is called a solid-state storage device, meaning there are no moving parts. It was developed by Toshiba from EEPROM (electrically erasable programmable read-only memory).

Flash memory has a grid of columns and rows with a cell that has two transistors at each intersection. The two transistors are separated by a thin layer of oxide. The floating gate’s only link to the row (wordline) is through the control gate. As long as the link is in place, the cell has a value of 1. To change it to 0, Fowler-Nordheim Tunneling must be done.

Tunneling is used to alter the placement of electrons in the floating gate. An electrical charge comes from the column (bitline), enters the floating gate and drains to the ground. This charge makes the floating-gate transistor act like an electron gun. The excited electrons are pushed through and trapped on the other side of the thin oxide layer, giving it a negative charge. These negatively charged electrons act as a barrier between the control gate and the floating gate. A special device called a cell sensor monitors the level of the charge passing through the floating gate. If the flow through the gate is above 50%, it has a value of 1. Otherwise it has a value of 0. By default all the gates in a blank EEPROM are fully open, giving each cell a value of 1.

The electrons in the cells can be returned to normal by applying an electric field, a higher voltage charge. Flash memory uses in-circuit wiring to apply such a field to either the entire chip, or predetermined sections called blocks. This erases those areas of the chip, and they can be rewritten. Flash memory is faster that EEPROM since it erases a block at a time instead of a byte at a time.

There are two main types of flash memory, NAND and NOR.

NOR flash was used initially as a replacement to ROM. Not allows data to be read bitwise, so it does not need to be loaded onto RAM. This means it can be, and indeed is, used as firmware. It has a fast read time, but a long write time. Thus, it is good for applications that do not need to write data, like BIOS. Writing repeatedly onto NOR flash may cause it to crash. For NOR flash, having more memory requires extra wiring, which means a larger physical size.

NAND flash has fast write speeds and normal read speeds. However, it cannot address data bitwise. It has a block-wise read and write mechanism, thus it requires RAM to be written onto. It cannot be used as firmware. It is physically smaller that NOR flash and is thus used as secondary hard disks, such as pen drives. NAND flash tends to be more expensive than NOR flash.

Hard disks are mechanical devices so they may fail. Flash memory is electrical, so it does not normally fail. This, along with their smaller size and faster speeds, is making flash devices increasing popular as a replacement to hard disks. Commonly, these replacements are called Solid State Drives (SSD). However, hard disks are still much cheaper than SSDs. There are also smaller flash devices like SD cards. They have small controllers attached to them. The technology used for these is different from the original invention by Toshiba.

## Floating-Gate Transistors

Flash memory stores information in an array of memory cells made from floating-gate transistors. Each memory cell is a standard metal-oxide-semiconductor field-effect transistor (MOSFET), except that the transistor has two gates instead of one. The cells can be seen as an electrical switch in which current flows between two terminals, and is controlled by a floating gate and a control gate. The floating gate is electrically isolated by an oxide layer, so electrons placed on it are trapped. When the floating gate is charged with electrons, this charge screens the electric field from the control gate, increasing the threshold voltage of the cell. This means that now, a higher voltage must be applied to the control gate to make the channel conductive. In order to read a value from the transistor, an intermediate voltage between the threshold voltages is applied to the control gate. If the channel conducts, the floating gate must be uncharged, otherwise, the floating gate is charged. The prior case gives a logic ‘1’, while the latter gives a logic ‘0’. The presence of a logic ‘1’ or ‘0’ can thus be detected. By default, every voltage is 1.

In flash memory, all bits on which data must be written, must be erased first, since they already contain electrons. 0 means there are some electrons in the floating gate, which must be sucked out to return the floating gate to the normal state.

USB devices used to have two power lines, one for USB and one to operate the device itself. Now they do not. They use charge pumps, to get higher voltages which are needed to write and erase data. Charge pumps waste half of the power. If the charge pump fails, data can only be read, not written.

NAND Flash puts the wiring in a series, so all of it must be called together (blocks). To call a certain bit is given a slightly higher volt.