



*Hybrid Memory Products Ltd*

# Floating Gate Memory Arrays

## Retention Issues

### Introduction

This document reviews the nature of Data Retention, and presents approaches to its specification and verification based upon procedures defined in the following IEEE publication:

***P1005 - Standard for Definitions, Symbols and Characteristics of Floating Gate Memory Arrays.***

### Data Retention

#### *Background*

Logic "ones" and "zeros" are represented in Flash or EEPROM memory cells by the amount of charge on the floating gate. Specifically, information is stored on the floating gate in the form of added or missing electrons. The time elapsed between the original transfer of electrons by the act of programming or erasing, and the first erroneous readout from that memory cell is called retention.

Stored charge is subject to:

1. Natural decay mechanisms
2. Decay or charging associated with material or geometry defects
3. Possible charge transfer effects associated with circuit design.

Retention encompasses all of these effects, and is a measure of the life of the stored information, or of the non-volatility of the data. In production worthy floating gate devices the intrinsic retention characteristics are millions of years.

Naturally occurring charge loss or charge gain is usually associated with:

- i. Gate oxide or interpoly dielectric defects, often observed as single random bit's failure,
- ii. Compensation of stored charge by ionic contamination, often observed as clustered bit's failure,
- iii. Releasing of holes from drain-avalanche induced hole traps, often seen as bit's failure,

- iv. Conduction of electrons through oxide or ONO on non-defective bits, a very slow leaking process over long period time.

With better understanding of floating gate property and non-volatile memory products (EPROM, EEPROM and Flash EPROM's), the first two kinds of charge loss or charge gain failures associated with defects and contamination are usually screened through voltage and temperature stress. The third kind of charge loss usually associated with the drain or source program/erase over voltage that can be carefully designed out. The fourth kind of charge leaking phenomena is associated with the oxide or ONO intrinsic property which has been measured to result in a loss of less than one electron per day over specified operation or storage conditions. This indicates that floating gate memories can retain data for millions of years at room temperature based upon extrapolations of the intrinsic charge loss mechanisms.

The ionic contamination can be observed by high temperature bake. In this case positive ions move towards floating gate due to the electric field and the diffusion gradient. Most of the literature has focused on positively charged alkali ions (e.g. Na<sup>+</sup> and K<sup>+</sup>) as the causative agents for ionic charge loss. However, H<sup>+</sup> is also a positively charged ion that is quite mobile in SiO<sub>2</sub> at temperatures at which floating gate devices are operated and stored. There is literature indicting H<sup>+</sup> as a active charge loss agent. The charge loss which has been reported to result from the ingress of moisture is probably also related to H<sup>+</sup> ions that are liberated by the reaction of the moisture with the SiO<sub>2</sub> lattice. Charge loss occurs when the floating gate loses electrons. This typically occurs on negatively charged or programmed floating gates. Charge gain occurs when the floating gate attracts electrons. This typically occurs on neutral or erased floating gates.

## Concept of Retention

### *Definition*

Retention time is the time from data storage to the time at which a verifiable error is detected from any cause. When an error is detected, the cause of failure must then be determined. Retention is a measure of the integrity of the stored data as a function of time.

For retention to have meaning in the context of a device specification the data should be stored using conditions within the allowed range of parameters. The data storage period should also be limited to the allowed parameter range.

Operating conditions during the storage and retention time periods may vary widely. Those conditions permitted by the data sheet that lead to the shortest retention time should be used in testing for retention.

The memory device under consideration may be a single transistor, it may be a memory cell consisting of several transistors, or it may be a large array of memory cells supported by peripheral circuitry to function as an EEPROM. In all cases, the first verifiable malfunction during readout defines the end of the retention of that particular device.

### *Memory Cell Retention*

Charge loss or gain may occur because of intrinsic mechanisms or because of defect related mechanisms. Floating gate memory cell can be designed so that the charge

transfer due to intrinsic mechanisms is minimal, and retention times for a defect free cell are so long as to be non-measurable.

For floating gate structures, the primary intrinsic charge transfer mechanisms are Fowler-Nordheim tunneling and Schottky-Richardson emission. A simple Fowler-Nordheim model and the Schottky-Richardson model both exhibit very low leakage currents and, hence, long retention times under device operating and storage conditions as is shown in the following paragraphs.

Consider an example where oxide insulation surrounds the floating gate. In one region the electric field is high enough to allow tunneling to occur; charge transport in other, lower field regions can be neglected. Even when all surrounding conductors are grounded, the potential resulting from the stored charges will cause a very slow charge transfer through the operation of the Fowler-Nordheim tunneling mechanisms.

For this simplified case, intrinsic charge transfer can be predicted from the expression derived in the appendix to this section, viz:

$$t_R = (C/a \cdot b) \exp(bV_t) \quad (\text{Eq 1})$$

where  $t_R$  is the retention time,  $C$  is the capacitance of the floating gate,  $V_t$  is the effective voltage resulting from the stored charge at which failure occurs, and  $a$  and  $b$  are constants of the Fowler-Nordheim equation.

When the pertinent and separately measurable values for all the quantities to the right of  $t_R$  are substituted, retention lifetimes in excess of millions of years result for well designed memory cells. This is true, whether the floating gate is part of a thin oxide cell, or whether the thicker oxide of the textured poly technology surrounds it. This large retention is an intrinsic property of floating gate memory technologies, and it can be expected from defect free memory cells.

At high temperatures, electrons will also be transferred by Schottky emission, which depends on an interplay of temperature, activation energy, and applied field. It is governed by the Schottky-Richardson equation, viz:

$$J = AT^2 \exp \left\{ -\sqrt{\phi - q^3 E / 4\pi\epsilon_0} / kT \right\} \quad (\text{Eq 2})$$

where  $J$  is the current density,  $A$  the Richardson constant,  $\phi$  is the emission barrier height,  $q$  is the electronic charge,  $E$  is the self-induced electric field, and other symbols have their usual meanings. This mechanism gives rise to a temperature dependent retention, and has measurable activation energy of 1.7 eV.

With intrinsic retention times exceeding millions of years in the operating temperature range, it is clear that real time retention data cannot be measured. Only under almost unrealistically high accelerating conditions at temperatures above 300 °C, have intrinsic retention's of the order of months been measured. This can be translated into a retention of 1 million years at 15°C, or about 120 million years at 56°C.

In rare and isolated instances, a defect will occur in a memory cell. Such a defect may give rise to physical constants that will result in a shorter retention for the affected cell. Activation energies of 0.6-0.8 eV have been measured for such defects. In principle, there is also the possibility that hot electrons injected from the substrate are collected inadvertently by a floating gate. This latter effect can be

eliminated by proper cell design. Finally, charge transfer as a result of charge compensation from impurities has been characterized with activation energies of 1.2 eV to 1.4 eV.

## Retention in integrated Circuits

Retention time for an integrated circuit is defined to be the retention time for the worst cell in the integrated circuit. This simple fact points out some of the relationships between memory cell and integrated circuit retention.

First, because retention is determined by the worst cell, it is statistically dependent on the number of the cells, and on the distribution of cell retention.

Second, a cell embedded in an integrated circuit can be expected to be exposed to a varied set of operating conditions. The details of those conditions will be specific to a given circuit and to a given fabrication technology. This makes it difficult to generalize as to what might constitute a worst case combination of conditions for test purposes.

Third, the presence of a very large number of cells increases the probability that one of them will contain a defect. In high density EEPROMs it can therefore be expected that retention (the failure of the first cell) will have a component that is determined by the defect density inherent in the technology. Production screens can eliminate structural defects giving rise to infant failures. The ultimate retention failure rate is caused by statistical defects inherent in a given oxide process.

## Specification of Retention

Stating a retention time requirement for a floating gate device is equivalent to saying that the defect related retention will not fall below that amount. The intrinsic retention time is so large that it is irrelevant.

Defect related retention is not a measurable parameter; it is a result of latent failures which become manifest. It can be best managed statistically, and can be specified as a failure rate established at some confidence level. Common practice is to determine the failure rate using temperature as an acceleration factor; the particular test temperature and the acceleration factor are parameters which must be defined.

The specification of retention is really a way of specifying the life expectancy of the data in a given product. Therefore the concern is with failures occurring early in life (infant mortality) and with the random failure rate during normal use.

Infant mortality should properly not be a concern of the end user. The device manufacturer should establish combinations of accelerating stresses and tests which detect and eliminate such defects before the product is shipped.

A representative specification for retention may be stated as: " 5 FITs (upper bound at the 60% confidence level) at 55°C ". This follows the practice established for (UV erasable) EPROMs, which is a floating gate memory technology with a large database. The term FIT comes from 'Failure in Time', and is defined as one failure per 10<sup>9</sup> device hours. Present practice is to use 55°C or 150°C for the reference temperature and 60% or 90% for the confidence level.

## Verification of Retention

### *Engineering Analysis Techniques*

The retention characteristics of a floating gate device may be examined by subjecting samples of the product to tests involving high temperature storage of the part during the retention time period. Such tests are time consuming and require care in device handling.

In the most general case, one might want to examine the defect related failure rate, the defect related activation energy, the intrinsic retention failure rate, and the intrinsic retention activation energy. During the life of a given technology, it is good practice to perform all of these experiments.

If one is willing to accept activation energies reported in the literature (or have otherwise predetermined them) and is interested only in whether the failure rate exceeds some goal, it is possible to obtain information from a simple experiment.

Suppose that the product specification is 5 FITs at 55°C with a 60% confidence level. The literature reports that the apparent activation energy for mechanisms involved in random, oxide defect related failure of floating gate devices is about 0.6 eV. Using a test temperature of 250°C, a sample of 75 parts would have to retain data for 1000 hours (6 weeks) with no failures in order to confirm the FIT specification. If one failure is to be permitted, the sample size must be doubled.

Similar experiments aimed at confirmation of the intrinsic retention values would require temperatures of 300°C, 325°C, and 340°C to verify mean intrinsic retention times of 3500, 900, and 330 hours respectively assuming an activation energy of 1.6 eV.

Working at high temperatures for any experiment of this type requires that consideration be given to package limitations. Ceramic packages are preferable, as plastic packages are limited to a temperature of 150°C. Retention experiments at 150°C require large sample sizes. For example to confirm the 5 FITs at 55°C as discussed above with plastic packages would require a sample of 7750 devices.

### *Production Screening Approaches*

The manufacturer should subject the part to stresses capable of accelerating infant mortality retention failures, and should conduct tests capable of detecting the failures. Individual manufacturers will have different approaches to accomplish this goal, and should document the efficacy of their particular choice.

One screening approach for potential infant mortality failures involves baking the devices with a data pattern opposite to that of the Equilibrium State at 150°C for 48 to 72 hours, or at 170°C for 12 to 24 hours followed by testing. This is generally enough to detect the infant failures.

## Attachment 1 - Data Retention Testing of HN58C1001

---

Proposal is as follows:

1. We set a target specification of 5 fits (60% Confidence Level) at 55°C operating temperature. This is an industry standard specification and equates to 5 failures maximum in  $10^9$  device operating hours.
2. In order to accumulate sufficient device hours to support this specification we would need to test 75+ devices (in ceramic package) for 1000 hours at 250°C
3. Prior to testing the 75 devices would be subjected to HMP standard non-volatile screen of 1000 read/write cycles followed by 72 hours retention bake at 150°C to detect early life failures.
4. Once testing is completed, calculations can be made to determine the expected fit rate at different temperatures or confidence levels as required.

The proposed test flow is summarized below:

