

# Reliability Considerations for Reprogrammable Nonvolatile Memories



Technical Paper  
November 2001

## INTRODUCTION

When acquiring a microcircuit, many considerations above and beyond the purchase price are important. Among these are quality, reliability, delivery, service, and product assurance. The lowest cost of ownership for the user is a result of the proper balance and specification of the above considerations. Reprogrammable nonvolatile memories, i.e., EEPROMs, contain the reliability considerations of endurance and data retention that can significantly affect the cost of ownership if the EEPROM is incorrectly used in the application.

## RELIABILITY

For EEPROMs, reliability is the summation of the factors of operating life (read), data retention, and endurance.

$$F.R._{Total} = F.R._{Read} + F.R._{Data\ Retention} + F.R._{Endurance}$$

F.R. is the failure rate. Read, Data Retention, and Endurance define the failure factor.

### Read Failure Rate

Read is the typical failure rate associated with dynamic high temperature life test. In general, EEPROMs have lower, i.e., better Read failure rates, because of the screening required to assure low data retention and endurance failure rates.

### Data Retention Failure Rate

Data Retention is the ability of the EEPROM to retain data. For most floating gate structures, the intrinsic data retention duration is so great, e.g., 100s to 1,000s of years, that for all practical purposes the intrinsic data retention failure rate is negligible. Extrinsic data retention duration is a function of endurance and may degrade with accumulated endurance cycles. The extrinsic data retention failure rate is included in the endurance failure rate.

### Endurance Failure Rate

Endurance is the most important failure factor because the endurance reliability is a direct function of the application, i.e., the number of times the device is rewritten during system operation. In other words, the total system life can be compromised by the endurance capability of the EEPROM. See Figure 1. The correct endurance value of the EEPROM must be selected to optimize the system performance. For applications not requiring many rewrites, e.g., BIOS program storage, EEPROMs with a low number of guaranteed endurance cycles provide the best combina-

tion of cost and reliability. For applications requiring many rewrites, e.g., data storage or configuration, EEPROMs with a high number of guaranteed endurance cycles provide the best combination of cost and reliability. For all applications, quality, delivery, service, and product assurance are identical for SST EEPROMs.

Endurance is defined as: "The measure of the ability of a nonvolatile memory device to meet its data sheet specifications as a function of accumulated nonvolatile data changes," per IEEE Std 1005-1998 *IEEE Standard Definitions and Characterization of Floating Gate Semiconductor Arrays*. The data sheet specifications include Write functionality, data retention, and Read access time. For an SST EEPROM, a nonvolatile data change is the completion of an Erase/Program cycle for each cell, i.e., transferring charge from and to the floating gate in the memory storage transistor.

All floating gate memory cells are subject to 9 possible endurance failure mode major categories (see Table 1). Endurance is unlike other MOS reliability concerns in that the device can be operated within the data sheet limits and eventually an endurance failure will occur. This is because the oxides through which the charge is transferred to and from the floating gate are nominally insulators; however, are subject to electrical stress from the Erase and Programming operations. The basic endurance failure mechanisms of oxide damage and charge trapping are caused by the cumulative effects of passing a current through the oxide and placing a high electric field across the oxide.

The thickness and integrity of the oxide are the primary factors in determining the endurance capability of the device. Thicker oxides have a greater likelihood of measurable charge trapping. Thinner oxides require greater care in processing to reduce defects; initial defects cause yield loss, latent defects cause endurance failures. Erase/ Program cycling over the lifetime of the system will cause random oxide damage and charge trapping at some constant low level. Design must be such to minimize the stress on the affected oxides; thus, reducing the generated defect and charge trapping rates. Processing must be such to minimize initial defects and reduce latent defects to the lowest possible level.

Although all floating gate memory cells are subject to the same failure modes, some of the failure modes can be eliminated through process architecture or logic design. SST has patented a memory cell design and associated SuperFlash fabrication process that is less subject to endurance failure modes than competing architectures. In all cases, stressing and testing must assure that defects



## Technical Paper

are identified prior to shipping. Table 1 summarizes the susceptibility of the major types of flash EEPROMs to the various endurance failure modes.

### ENDURANCE FAILURE MODES:

The failure modes are defined as follows:

#### Stuck Bit

A bit is unable to change and can be stuck in either logic state. A stuck bit can be caused either by charge trapping or oxide rupture. If caused by charge trapping, the trapped charge may disperse after some time (which can be accelerated by temperature) and the device will regain functionality. If caused by oxide ruptures, the cell will go to its intrinsic state, i.e., no charge on the floating gate. Sometimes referred to as “fast leaky bits,” stuck bits are verified by extended endurance cycling.

#### Retention Degradation

The loss of charge of the floating gate caused by either trapped charge or damage in the insulating oxide. Sometimes referred to as “slow leaky bits.” This extrinsic data retention duration is verified by a data retention bake after the specified number of endurance cycles.

#### Read Time Degradation

The gradual increase in the read access time caused by accumulated trapped charge or gradual charge loss reducing the cell current. Verified by worst case speed testing after endurance and dynamic high temperature life stressing.

#### Erase Time Degradation

The gradual increase in the time required to erase the memory caused by accumulated trapped charge. Degradation effects may be avoided by proper guardbanding of erase times or use of iterative erase algorithms with an erase threshold reference circuit.

#### Program Time Degradation

The gradual increase in the time required to program the memory caused by accumulated trapped charge. Degradation effects may be avoided by proper guardbanding of Program times or use of iterative program algorithms with a program threshold reference circuit.

#### Disturbs

Disturbs are an intrinsic phenomena of all memory arrays. A disturb occurs when reading, erasing, or programming one location causes an unwanted alteration at another location. Floating gate arrays require special considerations in design and processing to minimize unwanted effects of the high voltage used for program or erase.

#### Overerase

A device is unable to Read or Program correctly because of excessive memory transistor source-drain current, which grounds the bit line read or programming voltage. Overerase is eliminated in two transistor or split-gate memory cells by the isolation of the memory transistor from the bit line. Overerase is caused by hole trapping in the oxide, which increases the normal variation in the erase threshold distribution of memory transistors in large arrays, both initially and after Erase/Program cycling. Overerase effects can not occur in SST devices because of the split-gate design.

#### Erase Disturb

Unintentionally changing the contents in a non-accessed location, while erasing another location. This occurs because the high voltage required to Erase may not be isolated from the non-accessed locations. Erase disturb is impossible to occur in SST devices because all locations subject to the erase high voltage are erased simultaneously.

#### Program Disturb

Unintentionally changing the contents in a non-accessed location, while programming another location. This occurs because the high voltage required to Program may not be isolated from the non-accessed locations. Program disturb is minimized by proper design and processing, defects are eliminated through screening.

#### Read Disturb

Unintentionally changing the contents in a location, while reading that location or another location. This occurs because the voltage required to read may affect other locations. Read disturb is minimized for SST devices by clamping or regulating the word line.



# Reliability Considerations for Reprogrammable Nonvolatile Memories

Technical Paper

**TABLE 1: ENDURANCE FAILURE MODES**

Failure Mode	Thin Oxide Stacked Gate	Thin Oxide Two Transistor	Thick Oxide Split-Gate (SST)
Stuck Bit	Lo <sup>1</sup>	Hi <sup>2</sup>	Lo
Retention Degradation	Hi	Hi	Lo
Read Time Degradation	Lo	Hi	Lo
Erase Time Degradation	Hi	Lo	Hi
Program Time Degradation	Hi	Lo	Lo
<b>Disturbs</b>			
Overerase	Hi	N/A <sup>3</sup>	N/A
Erase	Hi	Lo	N/A
Program	Lo	Lo	Lo
Read	Lo	Lo	Lo

T1.1 704

1. "Lo" means the technology is not comparatively susceptible to the identified failure mode.
2. "Hi" means the technology is comparatively susceptible to the identified failure mode.
3. "N/A" means the failure mode is "Not Applicable" to that technology.

## SUMMARY

An extremely important benefit of the SST SuperFlash EEPROM is the complete elimination of the possibility of "overerase" effects by use of the split-gate isolation and "erase disturb" by use of the word line erase feature.

Endurance follows the "bathtub" curve, with a known infant mortality region, a useful life region, and a predictable wear-out region. Endurance cycling for screening and for periodic qualification testing has historically been considered the preferred means of verifying capability. Cycling can be performed in real time and may use the actual operating mode of the device or a test mode. Cycling is often combined with voltage or temperature stressing to accelerate infant mortality failures. A significant disadvantage of stacked gate memory devices is the long time required to erase/program; thus, the time to gather reliability and process improvement data is greatly extended.

Although intrinsic data retention duration is essentially infinity, the extrinsic data retention duration is a function of endurance. The endurance failure rate contains the extrinsic data retention failure rate induced by endurance. The intrinsic data retention failure rate is reported independent of endurance.

JEDEC Test Method A117 describes the procedures to be used when performing endurance cycling for screening or endurance performance verification. Various means exist to eliminate infant mortality, which will be a function of product design and the dominant failure mechanism of the process. For example, all EEPROM manufacturing flows contain an infant mortality data retention screen, which is typically an unbiased bake.

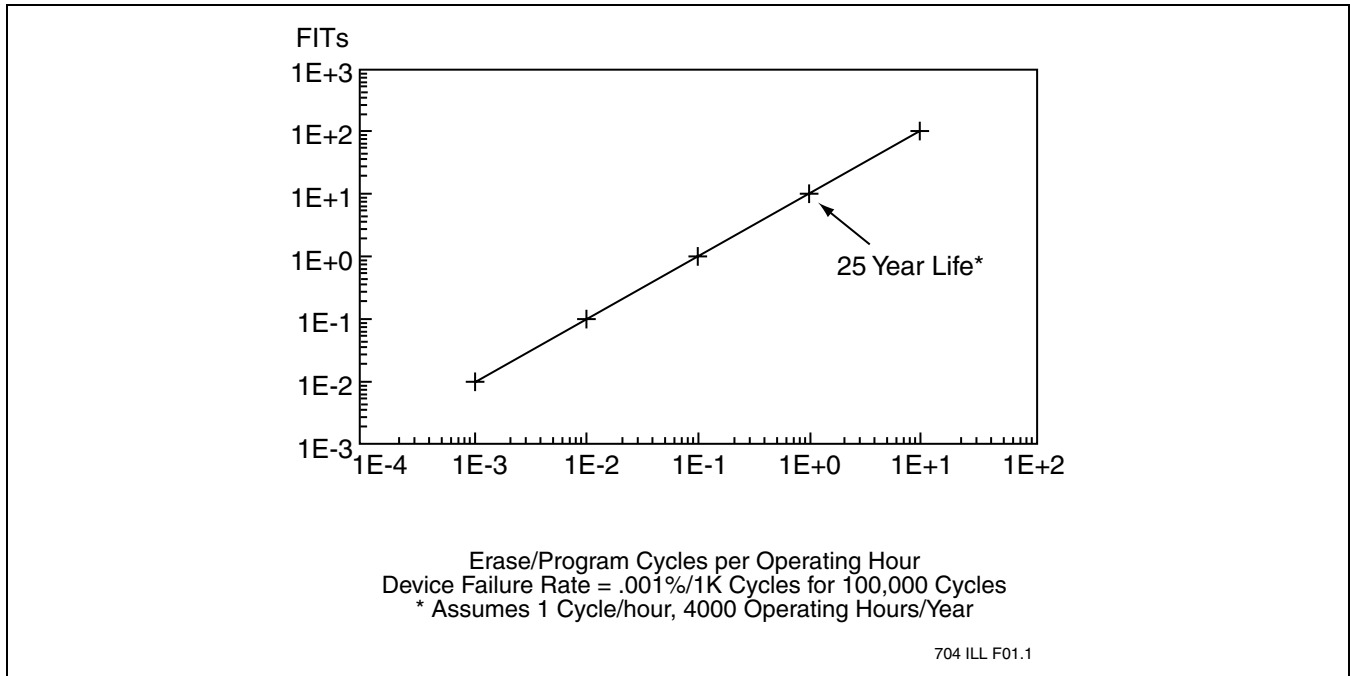
Random defects, occurring naturally in the wafer fabrication process, will cause infant mortality endurance or data retention failures. An explicit relationship does not exist that correlates infant mortality with performance in the useful life or wear-out regions. Extra memory, called redundancy, improves yields by repairing infant mortality failures in large memory arrays. The localized nature of the random defects causing infant mortality failures, assures the reliability of repaired and non-repaired devices is equivalent.

The endurance failure rate of the EEPROMs in a system will increase in importance as a function of the number of times the system rewrites the EEPROM during system life. System reliability is a function of the failure rate in the specified useful life region of the device, not when the onset of wear-out occurs. Given the operating life failure rate of a typical MOS memory is in the order of 100 FITs (.01%/1000 hours), the endurance failure rate contribution should be more than an order of magnitude lower, see Figure 1. SST EEPROMs will meet system reliability requirements by providing the lowest endurance and total device failure rate for the specified system lifetime. The greater system reliability lowers the cost of ownership of a SST EEPROM.



# Reliability Considerations for Reprogrammable Nonvolatile Memories

Technical Paper



**FIGURE 1: CONVERTING THE ENDURANCE FAILURE RATE IN %/1K CYCLES TO A FAILURE RATE IN TIME (FITs)**

- Note:** 1. Failures in Time = (E/P cycles/hr x End F.R. fails/cycle)/10<sup>-9</sup> hrs/FIT  
2. 25 years = (100,000 cycles)/(1 cycle/hr x 4,000 hours/year)