NOR Flash – Data Retention and Endurance

Introduction

Data retention and endurance are critical features of a NOR flash device. Every NOR flash device must meet certain data retention and endurance specifications. These specifications guarantee long service life.

This application note explains the physical mechanisms of NOR Flash device failure, the related testing procedures, and how to improve data retention and endurance.

The first part of this application note is more theoretical, providing background for understanding the issue. The later sections describe practical scenarios that are of interest to most customers.

Definitions:

- Data Retention
  - Data retention is the time (since last write) a NOR Flash device can reliably store data. This time is typically 20 years for Renesas products. This number is specified under defined conditions of wear and temperature for the device.
- Endurance
  - Flash memory endurance is a measure of how many times a block of flash memory can be reliably programmed and erased before it becomes unreliable.
  - All Renesas NOR Flash devices support 100,000 program-erase cycles with the exception a single product family.
  - This endurance value stated in our datasheet is a guaranteed minimum number of program-erase cycles that the device can endure while performing to datasheet requirements which were developed to be consistent with JEDEC standards.

Contents

1. Flash Device Operation 3

1.1 Flash Device Structure 3

1.2 Program 4

1.3 Erase 5

1.4 Oxide Damage 6

2. Endurance 7

3. Data Retention 7

4. Testing for Endurance and Data Retention 9

4.1 Program Erase Endurance Testing 9

4.2 Cycling Distribution Across Memory Blocks 9

4.3 Relaxation Methods 10

4.4 Data Retention Testing 11

4.5 High Jump Analogy 12

5. Improving Endurance and Data Retention 13

5.1 Endurance 13

5.2 Data Retention 13

5.3 Wear Leveling 13

5.4 Error Correction Codes 14

6. Summary 15

7. References 15

8. Revision History 15

A. The Physics of Oxide Degradation 16

B. The Arrhenius Equation 19

1. Flash Device Operation

1.1 Flash Device Structure

Figure 1 shows a basic example of a bit cell in a NOR Flash device. This N-channel transistor is turned on when the floating gate voltage becomes greater than the voltage threshold for the transistor.

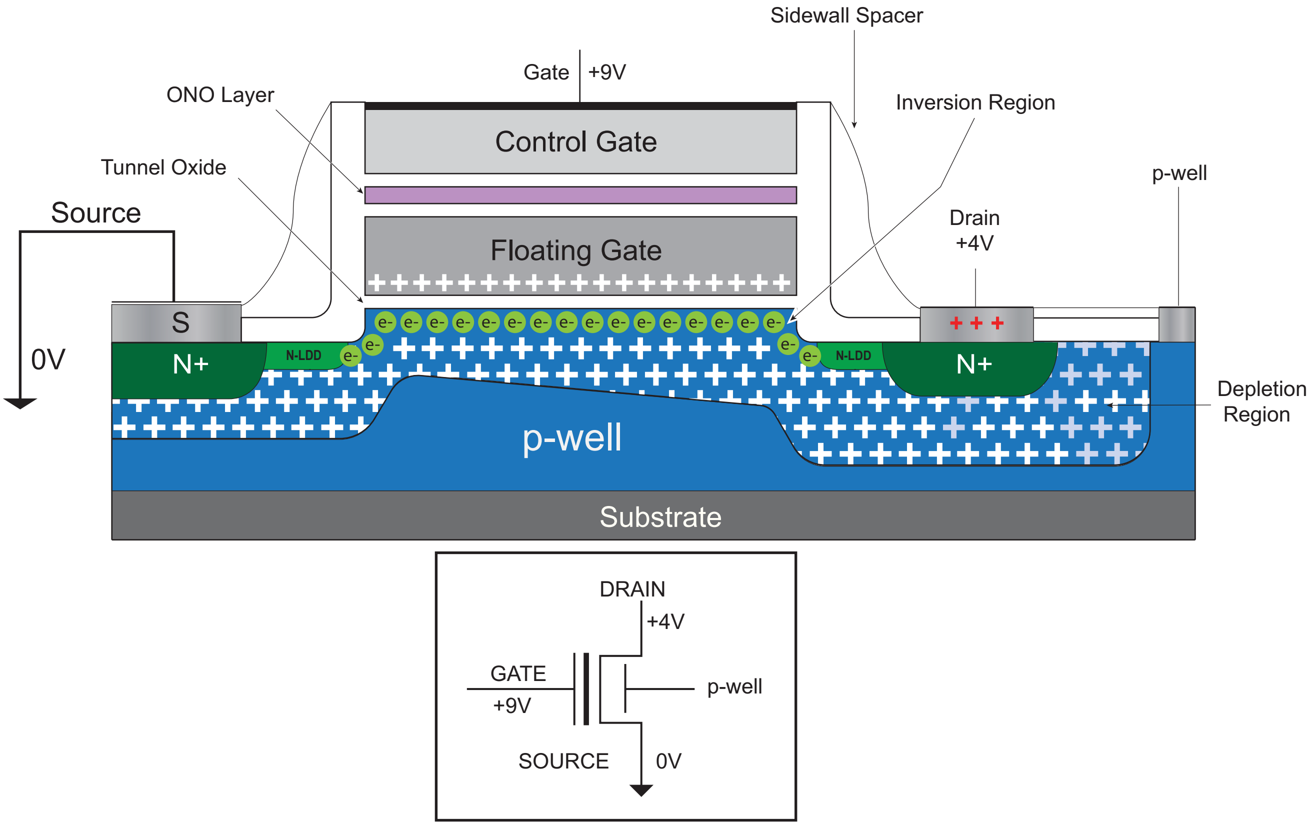


Figure 1. Cross Section of an N-channel Flash Device

1.2 Program

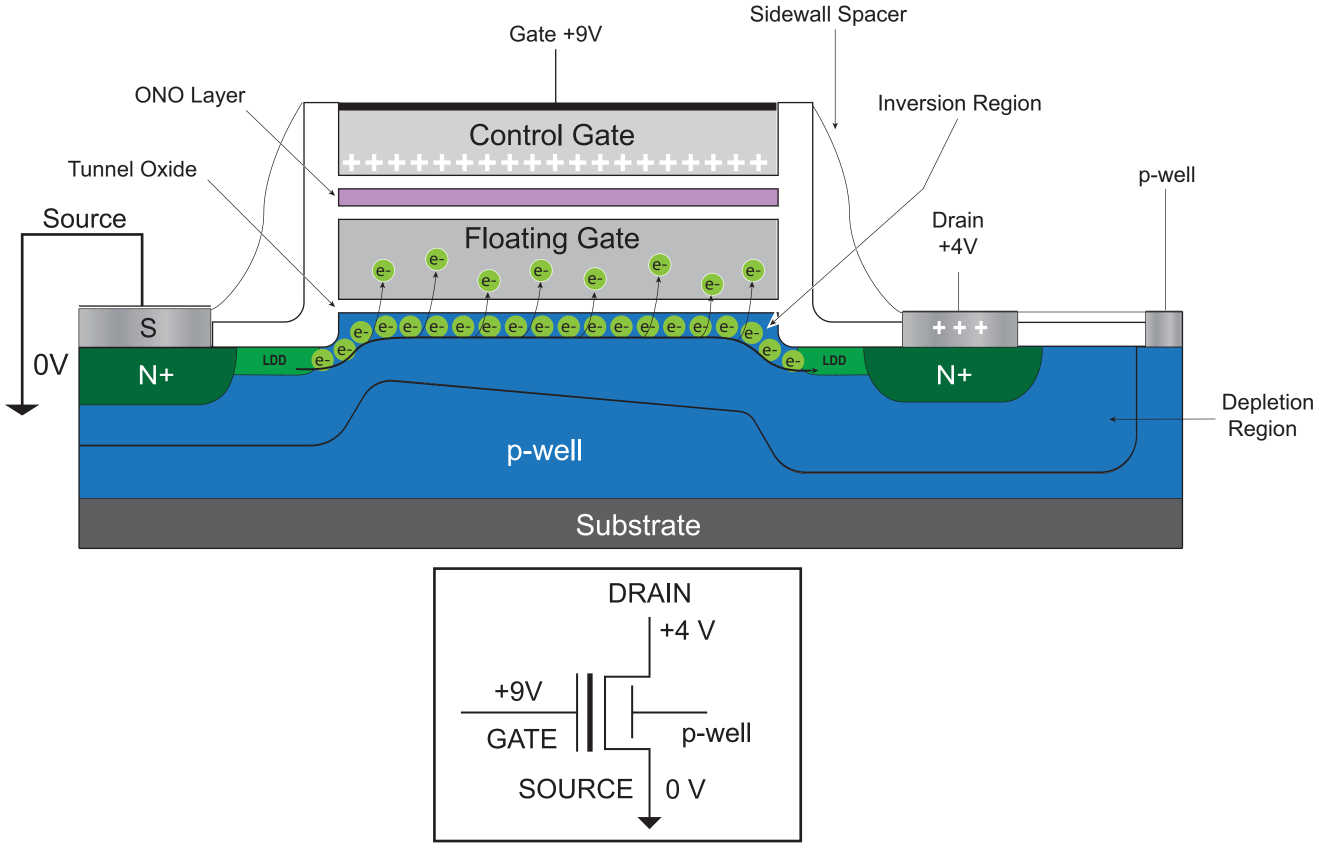


Figure 2. Cross Section of an N-channel Flash Device During Program

Programming (resulting in a logic 0 state) is achieved by driving electrons across from the source to the drain (as shown by the black arrows in Figure 2). These hot electrons interact with the atoms in the silicon and scatter into the oxide and across to the floating gate. The high positive voltage on the gate effectively "pulls" electrons into the floating gate. The larger the concentration of electrons inside the floating gate, the stronger the cell is programmed. This turns off the transistor current flow.

The Gate-Source Threshold voltage (V_{GSTH}) is the voltage on the control gate required for the MOSFET to conduct current. The higher the concentration of electrons inside the floating gate, the higher the applied Gate-Source Threshold voltage must be to overcome the negative bias voltage. That negative bias voltage is created by the trapped electrons inside the floating gate and is largely determined by how “strongly” (with more electrons), the MOSFET was programmed.

1.3 Erase

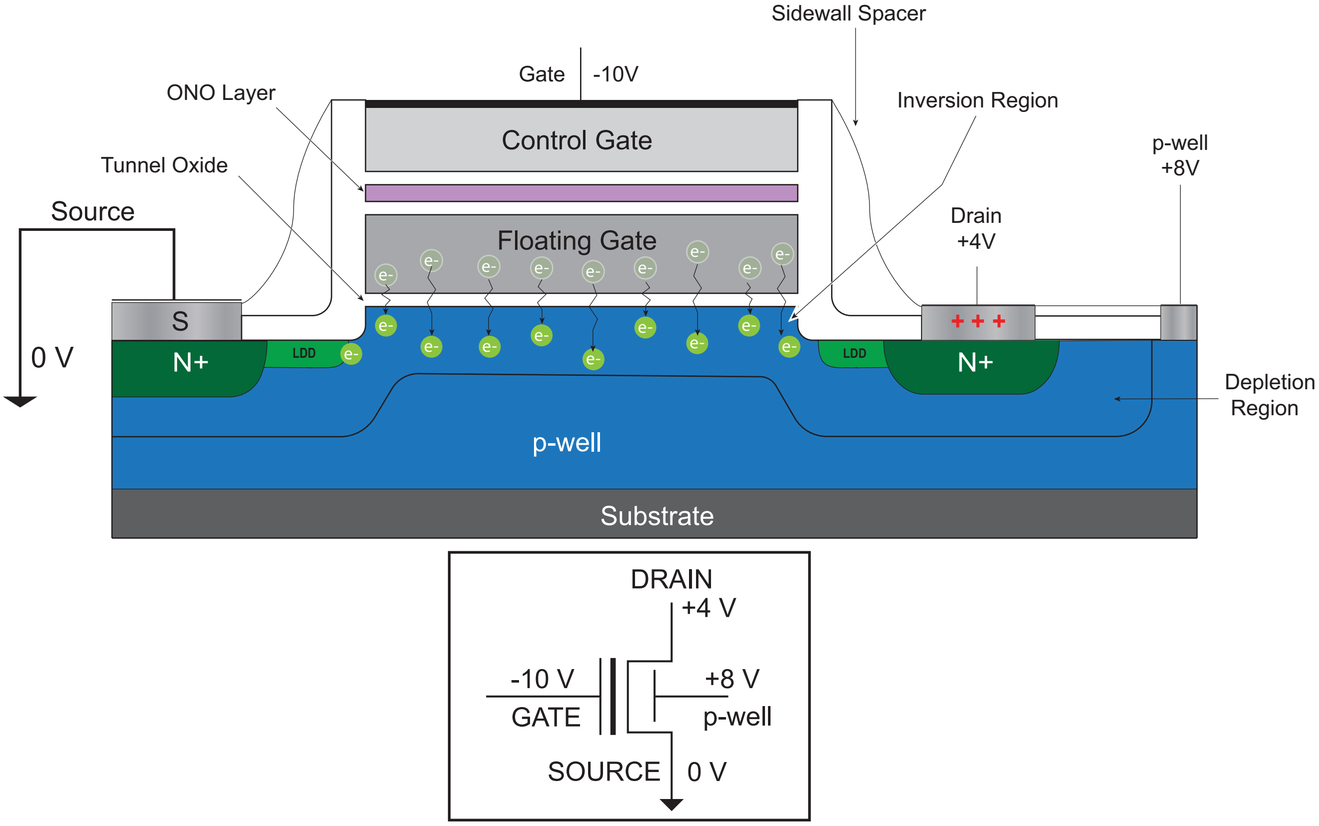


Figure 3. Cross Section of an N-channel Flash Device During Erase

For the erase function (resulting in a logic 1 state) the negative applied voltages from the control gate and the positive voltage on the p-well create a large electrical field that enables electrons to “tunnel” through the oxide barrier. The large electric potential difference between the gate and the substrate effectively “pulls out” electrons from the floating gate, via a mechanism called quantum tunneling. Quantum tunneling is the process of the electron passing through the oxide energy barrier when it does not have the energy to pass over the barrier height in the classic mechanical sense (see The Physics of Oxide Degradation). Through this process, the floating gate is depleted of electrons, lowering the effective Gate-Source Threshold voltage.

1.4 Oxide Damage

A major cause of degradation in NOR Flash devices is oxide damage caused by the flow of current through the oxide. With continued use, through repeated program and erase cycles, the oxide degrades to a point where the transistor may malfunction.

Over time, the tunneling oxide between the floating gate and the substrate changes. This happens as charges become trapped in the oxide. The traps in the oxide are initially caused by the manufacturing process from broken bonds in the oxide. This problem is partially healed by a forming gas anneal which completes some broken bonds with a Si-H bond. Over time, current causes an increase in the number of trapping sites (broken silicon-oxide bonds, and other bonds, Si-H, etc.) which allow even more charges to build up in the oxide.

Additional oxide damage is caused by holes that can be visualized as the lack of electrons and are effectively positive charges moving in the opposite direction to the electrons. Hole damage is a bit more complicated. The basic idea is explained by electron damage and therefore hole damage will not be covered.

For simplicity, the damage effects are illustrated using only electrons.

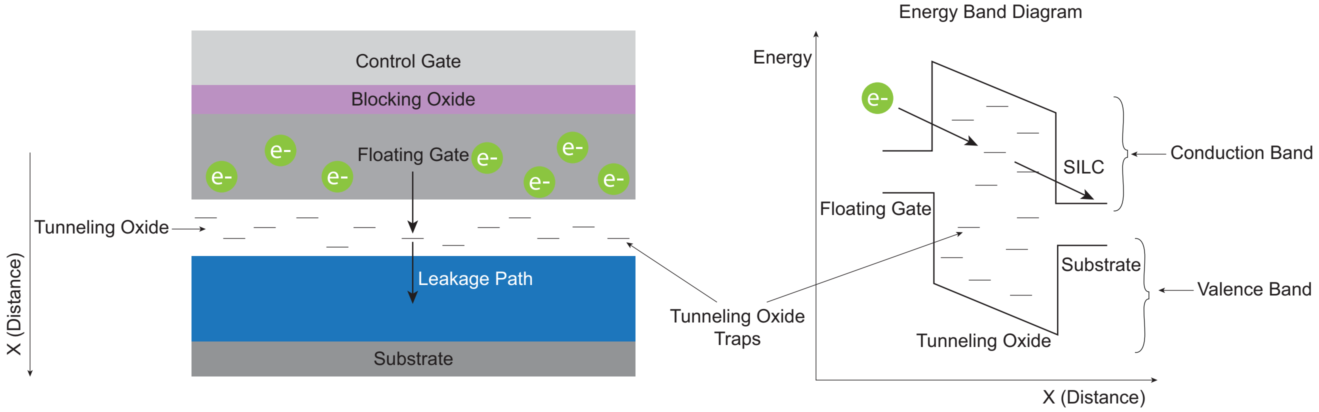


Figure 4. Close-up of the Tunneling Oxide and Trapping Band Diagram

Figure 4 illustrates negative charges in the oxide layer. The energy diagram on the right is a function of energy vs. distance. It shows the conduction band above and the valence band below with the traps distributed in energy value. SILC (Stress Induced Leakage Current) is the leakage from the floating gate to the substrate through the tunneling oxide.

If the degradation continues, the damage caused by the tunneling current will induce charge trapping sites. These sites allow trapping/detrapping in the tunneling oxide. Trapping is the capture of a charge in a trapping site. In the case of an electron, it is a negative charge. This trapping can dominate the transistor behavior by shielding the gate voltage from the conducting channel. The result can be a shift in both the threshold voltage and current output causing device failure.

Also, if continued, this damage can result in a complete breakdown of the oxide and shorting between the gate and substrate. An expanded discussion of the mechanism of this degradation can be found in the Wear Leveling Wikipedia article.

2. Endurance

The main causes of NOR Flash degradation after program/erase cycles include:

- Due to the elevated electric field that forces high-energy electrons to tunnel through thin oxide (~ 10 nm oxide for our devices), the tunnel oxide will degrade over time.
- Oxide degradation creates traps or states in the oxide, trapping electrons that shield the effective voltage on the floating gate. As a result, significantly more pulses are needed to program/erase, causing longer erase times after substantial cycling.
- Also, below the Gate Source Threshold, increased current causes a bit-line leakage that reduces the bit-line drain voltage during program operation which will slow down the programming time.
- The program/erase time will continue to increase until it exceeds the datasheet maximum specification.

Note: The maximum program/erase times specified in the datasheet are typically set to be the longest time measured (across all temperatures) when the Flash device reaches the endurance limit.

This leakage/shorting will limit endurance as the transistor fails to perform read, program and erase functions.

Degradation rate is calculated by an Arrhenius equation. That rate depends on the temperature the device experiences during cycling. For example, at high temperatures the degradation is accelerated. For more information, see The Arrhenius Equation.

3. Data Retention

At the NOR Flash cell level, data retention is the number of years from the last write until the floating gate charge is depleted causing an effective shift in the on/off state of the transistor. This may alter the value read from the cell (logic 0 or logic 1). This number of years is defined under specific conditions of wear and temperature of the storage device.

In floating-gate memories failure may be caused by charge trapping (normally accelerated by lower temperatures and/or shorter cycling delay) or detrapping (normally accelerated by higher temperatures) in the tunneling dielectric. Additionally, failure can be caused by oxide damage (normally accelerated by higher temperatures) in the tunneling dielectric.

There are two separate stages to the leakage process.

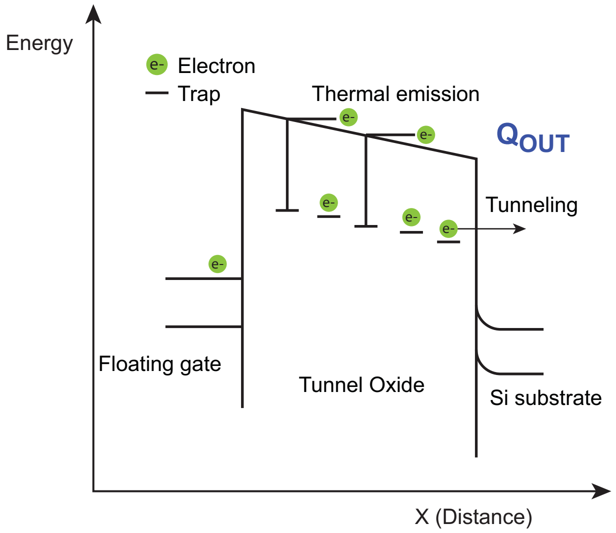
- Oxide wear-out which will eventually lead to high leakage.
- Charge loss from the floating gate during the data retention period.

For a greater number of program and erase cycles, the charge loss rate increases. The effect of oxide damage in the section above is the reduction in the time the floating gate can hold an effective charge, causing a change in floating gate voltage. For example, the transistor turns on at a lower gate source voltage than when it was originally programed. The cell read value may change from programmed (0) to erased (1).

At higher temperatures the electrons have more energy to escape the floating gate and so they leave at a faster rate. This shortens the period of the data retention. The 20-year period in the datasheet is specified for an operational temperature of 85°C or lower.

In a NOR Flash device the data is stored in the floating gate (FG) of a transistor. Inside the FG the electrons have some thermal energy for a given temperature. This energy allows the electrons to overcome the energy barrier of the oxide and leak to the substrate. The higher the temperature the greater this leakage, and the faster the reduction in the charge of the programming. In the case of a programmed logic 0 state the bit can flip to an erased state or a logic 1.

In addition, while in the 0 state, charge detrapping of electrons in the tunneling oxide can change the effective field of the floating gate charge. After the electrons have left the oxide, this will decrease the effective negative charge in the oxide and the negative field experienced by the channel. So, the transistor will not be fully turned off and will leak. In addition to the floating gate charge leakage, this reduced oxide charge will increase the probability that the transistor will turn on changing the transistor state from the programmed 0 to a 1.



- Thermal emission or tunneling (for traps close to the interface)
- High thermal activation (Ea = 1.1-1.3 eV)
- Logarithmic in time
- Slightly electric field dependent
- Accumulated charge dependent
- $Q_{OUT} = f(T, \log t, V_T, \text{Number of cycles})$
- T = Temperature in K
- t = Time

Figure 5. Detrapping: Flash Cells Recover After Relaxation or Bake

In a NOR Flash device the data is stored in the floating gate (FG) of a transistor. Inside the FG the stored electrons have some thermal energy allowing a few electrons to escape through the tunneling oxide. Statistically these escapes will increase with higher temperature.

Over time, the number of electrons decreases in the FG and after 20 years it is probable that some cells will have lost a substantial number of charges. This becomes a problem when the cell value either changes state from programmed to erased or the read value becomes unreliable.

It is highly likely that most of the cells are still OK (even with fewer electrons), but if even a single cell changes state, the stored data is considered corrupt. A flipped bit in Flash memory could cause a system malfunction due to execution of an incorrect CPU instruction or reading incorrect data.

If all cells were not programmed at the same time, the cells that were programmed first will reach the 20-year data retention time first. It is probable that these cells will have lost more of their trapped electrons and statistically are more likely to be the earliest to change their state.

Cycling endurance and data retention are not independent of one another; they are interrelated because the stress conditions of the oxide are common, and both are functions of temperature. This dependence is discussed in Testing for Endurance and Data Retention.

4. Testing for Endurance and Data Retention

Endurance and data retention testing are critical to assure the reliability of Nor Flash devices. NOR Flash vendors follow qualification guidelines published by JEDEC. Among other qualification procedures, vendors are required to perform endurance and data retention stress tests that comply with these specific JEDEC standards:

- JESD47: Stress-Test-Driven Qualification of Integrated Circuits
- JESD22-A117E: Electrically Erasable Programmable ROM (EEPROM) Program / Erase Endurance and Data Retention Stress Test (note: the term EEPROM includes flash memory)

These standards provide a comprehensive framework for semiconductor device qualification. They sample sizes, test structures, and pass/fail criteria for ensuring robust reliability. The tests simulate long-term field use conditions through accelerated stress testing.

The standards require that the flash devices undergo program/erase endurance cycling followed by retention testing. After passing these stress tests for a particular NOR flash product, a vendor can claim a minimum number of endurance cycles (typically 100K for Renesas) and a minimal data retention period (typically 20 years for Renesas) for that product.

4.1 Program Erase Endurance Testing

The non-volatile memory cycling endurance test (NVCE) validates that a NOR Flash device can sustain program and erase operations without failure up to its rated specification. A failure is defined as the inability to correctly program or erase cells or meet datasheet performance parameters such as maximum block erase time.

For the NVCE test, devices are split into two groups: one cycled at room temperature (25°C) and the other at elevated temperature (typically 85°C or higher).

Some guidance is provided regarding data patterns to be used in the programming part of the cycling.

The NVCE test time is limited to 500 hours.

JEDEC provides certain flexibility on NVCE test requirements which should be discussed here, specifically:

- Not all memory blocks must be cycled up to the limit (see cycling distribution below).
- Relaxation methods such as intentional delays and bakes for charge detrapping are allowed.

4.2 Cycling Distribution Across Memory Blocks

JEDEC recognizes that during product qualification it can be impractical to cycle every memory cell to the maximum program/erase (P/E) cycle limit, especially for large memories. To balance test realism, time constraints, and device qualification needs, the standard provides guidance on when to cycle the full memory array versus only a fraction of the memory blocks.

The JESD22-A117 standard allows tiered cycling, where different fractions of the memory are stressed to different limits. For example:

- 1% of the blocks to 100% max cycles
- 10% of the blocks to 50% max cycles
- Remaining blocks cycled to under 10% of max cycles or uncycled

The JESD47 standard provides the following guidance: at least one-third of the operations should be devoted to cycling blocks to 100% of maximum specification, if possible, within the specified cycling time frame. And at least one block of each Flash device must be cycled to the maximum cycle count, regardless of the time required.

At Renesas, where possible for products with smaller memory arrays, all blocks of a NOR flash device are cycled up to the limit, otherwise a tiered cycling approach is used.

Figure 6 illustrates distribution of memory cycling across the memory block of a flash device when full cycling, up to the specified limit, is not possible for all memory blocks.

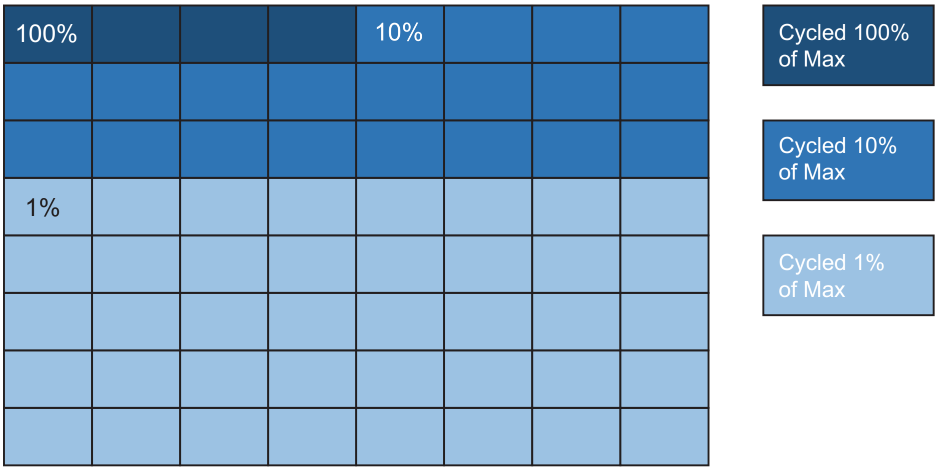


Figure 6. Tiered Cycling Distribution Across Memory Blocks

4.3 Relaxation Methods

JEDEC recognizes that, in real-world applications, non-volatile memories are not cycled continuously at high speed. Instead, program/erase (P/E) operations are spread over years of use. If qualification tests were to cycle devices continuously without breaks, the accelerated stress could cause unrealistic early failures, especially for certain recoverable degradation mechanisms.

To address this, the standard allows for intentional delays or other methods to simulate realistic usage patterns. These include:

1. Elevated Temperature Cycling with Built-In Relaxation
   a. Cycling is done at a higher temperature, with recovery time evenly distributed between individual cycles.
   b. Temperature acceleration is used to shorten the test time.
2. Reduced Cycling Frequency at Elevated Temperature
   a. Delays are inserted between individual cycles or groups of cycles to simulate the time that would normally elapse in the field.
   b. Cycling may be limited to a certain number of cycles per day, with the device idle the rest of the time.
3. Room-Temperature Cycling with High-Temperature Bake Intervals
   a. Cycling occurs at or near room temperature.
   b. At certain points, devices are baked at high temperature to simulate recovery that would normally happen at lower temperatures over a longer period.

Figure 7 shows an example of a cycling sequence for a particular memory block or group of blocks. In this example cycles are applied to the block in groups of 10,000 continuous cycles with a relaxation period of 10 hours between two groups of cycles.

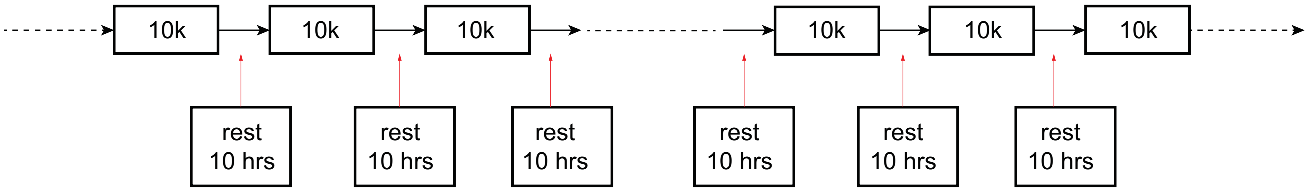


Figure 7. Memory Block Cycling Sequence

Essentially, JESD22-A117E provides structured methods to slow down endurance cycling or insert controlled pauses so that the cumulative damage to the memory device accurately represents years of field operation, rather than the unrealistic stress of continuous high-speed cycling.

4.4 Data Retention Testing

Data retention testing ensures that a NOR Flash device can reliably retain stored data for its expected lifetime, typically 20 years. The tests are applied both to fresh, uncycled devices as well as devices that have gone through extensive program/erase cycling. Data retention testing is a key element of device qualification for commercial and industrial applications.

Data retention behavior is typically simulated using an accelerated stress test at higher temperatures. The Arrhenius equation is used to relate test conditions to expected field lifetime. For example, an accelerated retention test at 125 °C may represent many years of use at 55°C. The activation energy value, which is used by the Arrhenius equation, is technology-dependent, with 1.1 eV commonly used for floating-gate devices such as NOR flash memories. For more information, see The Arrhenius Equation in Appendix A.

There are three main retention test conditions:

1. Uncycled High-Temperature Data Retention (UCHTDR): Tests unprogrammed devices to establish baseline retention behavior.
2. Post-Cycling High-Temperature Data Retention (PCHTDR): Evaluates retention after devices have undergone endurance cycling to simulate wear.
3. Low-Temperature Data Retention and Read Disturb (LTDDR): Detects failure mechanisms not accelerated by temperature, such as read disturb effects.

Acceptance Criteria:

- Zero failures are allowed within the tested sample groups for qualification.
- The resulting equivalent lifetime must meet or exceed the product’s expected service life.

The following chart illustrates how accelerated stress tests at elevated temperatures help in predicting the data retention period at a lower usage temperature by plotting the Arrhenius equation. The curves follow the Arrhenius equation across temperature. On the right-hand side of each curve, each data point represents a stress test condition (temperature and hours). Every point on the curve is essentially an extrapolation from the marked stress test point. So, given that the stress test passed, every usage condition represented by a point on the same curve is valid. The curves effectively represent the height of the bar the jumper must clear in the High Jump Analogy. On the left-hand side of each curve some equivalent usage conditions are marked as examples.

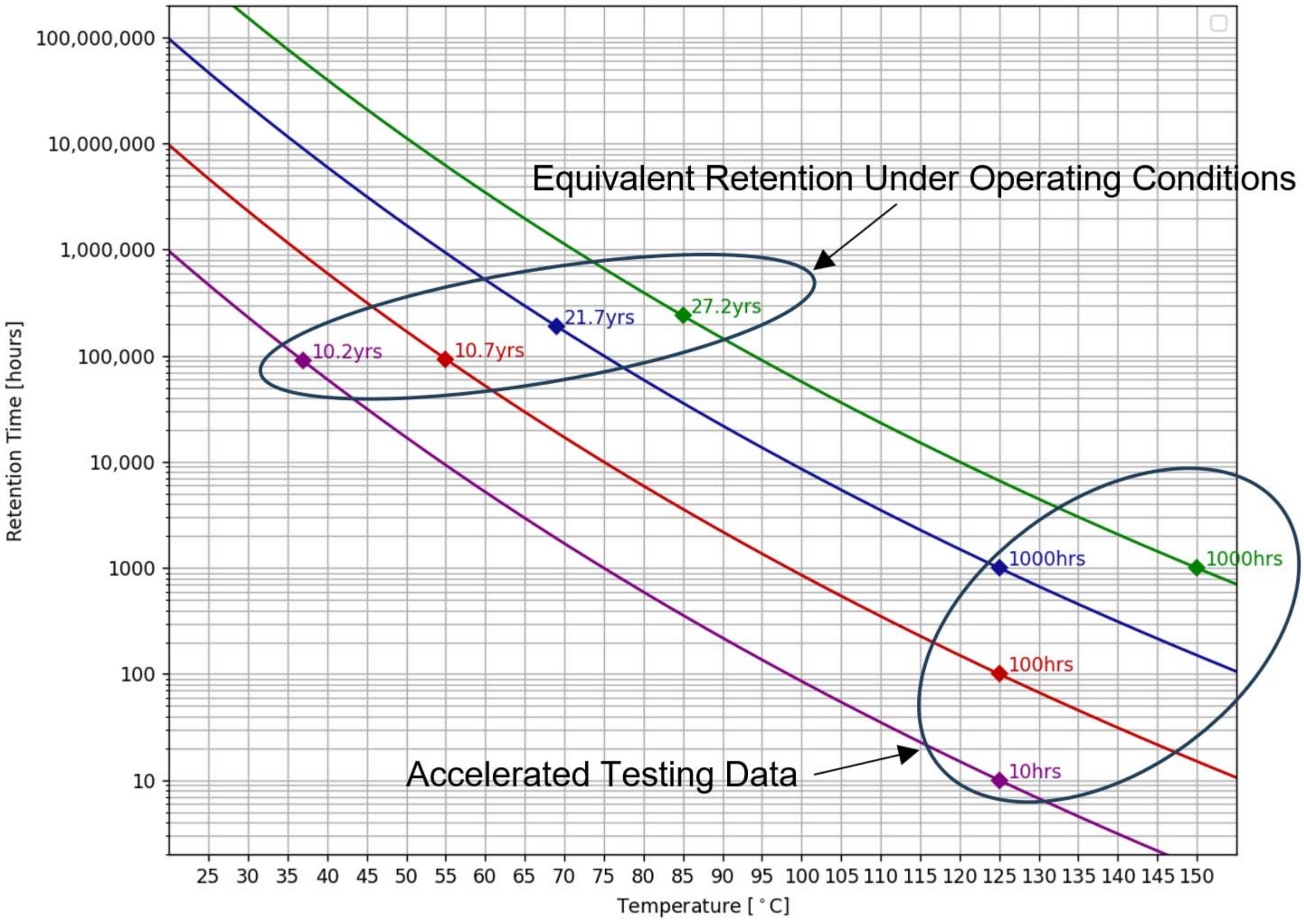


Figure 8. Accelerated Testing vs. Usage Operating Conditions

The stress test conditions of the top two curves (green and blue) are typically used in UCHTDR. That is uncycled preconditioning tests. The blue line is an uncycled preconditioning test well below the actual failure point from the 125°C testing point. The green line is generated from 150°C testing and is more aggressive, closer to the actual failure point. The stress test conditions of the bottom two curves are typically used for PCHTDR for devices cycled 10K times and 100K times respectively.

If we examine the top curve (green), the stress test condition is 1000 hours at a temperature of 150°C. On the left-hand side, we highlight an equivalent usage condition of 27.2 years at 85°C. The blue curve shows a stress condition of 1000h/125°C with a highlighted usage condition of 21.7y/69°C.

The red curve shows stress test condition of 100h/125°C equivalent to 10.7y/55°C, and the purple one shows stress test condition of 10h/125°C equivalent to 10.2y/37°C.

For more information on generating the data above, see The Arrhenius Equation.

4.5 High Jump Analogy

It is important to understand that if a device was qualified for a certain endurance limit or a data retention limit, this does not mean that the device cannot perform better than those limits. It just means that it passed certain tests with specific thresholds. This is similar to an athlete who participates in a high-jump competition and is rewarded based on the height of the bar that he/she has cleared. In reality, the athlete’s body may reach a height which is a few inches more than the bar’s height. Yet the athlete’s score will be based only on the height of the bar. Similarly, a NOR flash product is qualified based on successfully passing a specific test value. The part may be able to perform better than the specified value (greater endurance or data retention times). However, the test is standardized to confirm the values stated in the data sheet.

5. Improving Endurance and Data Retention

Given endurance and data retention limits, here are some ways to increase the service life of a NOR Flash device in a system.

5.1 Endurance

- If the device junction temperature can be lowered, the endurance will increase. This can be achieved by increasing system airflow, increasing thermal conduction of the board, or by heat sinking of the package.
- Endurance is also improved if the duty cycle of the program and erase function can be modified to allow additional relaxation time. This detrapping of electrons in the oxide removes the additional negative oxide charges that affect the floating gate programmed state.
- Wear Leveling – Distributing the cycling among as many blocks as possible ensures no block endures an excess amount of P/E cycles.

5.2 Data Retention

- The customer may consider data refreshing after a certain period. This can be done repeatedly during the product's lifetime. A balance between increased program/erase cycles and safe refresh times will reduce the probability of data corruption.
- Like the endurance suggestion, if the conditions can be improved by lowering use temperature through additional airflow, thermal conduction of the board or heat sinking of the package, the data retention time can be increased.

Refreshing the data well before the normal retention as specified (for example every few years) will improve the chances of retaining good data integrity.

5.3 Wear Leveling

Because there is a limit to the number of erase/program cycles that can be applied to each memory block, it makes sense to spread erase/program cycles across the memory blocks as much as possible. This is a technique known in the non-volatile memory world as wear leveling, and is intended to prolong the flash memory's service life.

At a high-level, wear leveling manages the distribution of erase/program operations across the memory array. It ensures that no single memory block prematurely fails by exceeding its endurance limit. Wear leveling spreads the program and erase cycles to as many memory blocks as possible, thus maximizing the number of successful cycles that can be performed during the system's lifetime.

Figure 9 and Figure 10 show simple illustrations of flash block cycling. The squares represent memory blocks, a blank cell represents a non-cycled block, and a blue cell represents a cycled block. The darker the blue color is, the more the block has been cycled. A red cell with 'x' represents a block that has exceeded its cycling limit.

Figure 9 represents a device where wear leveling is not implemented. Four memory blocks are cycled repeatedly, and finally all of them exceed their cycling limit before most available memory blocks have been touched.

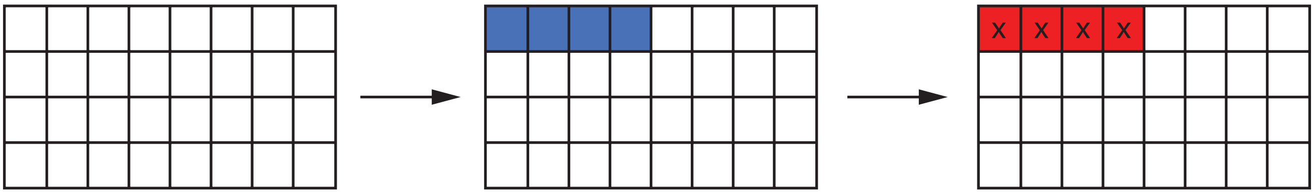


Figure 9. Device Without Wear Leveling Implemented

Figure 10 represents a device where wear leveling is implemented. Cycling is distributed between all memory blocks. The block cycling level increases uniformly across the memory array.

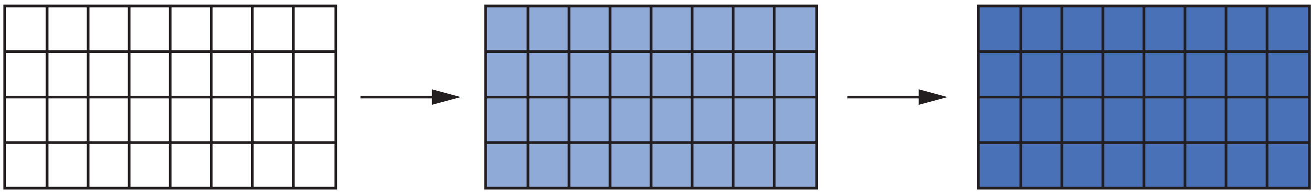


Figure 10. Device With Wear Leveling Implemented

The illustrations demonstrate the principle of wear leveling. Of course, the scenario can be more complicated. For example, half of the memory could be used for data logging. The other half could have a relatively stable image of code or data. At some point in the system's life, it would make sense to swap the two halves so that the frequently cycled region moves into memory blocks that have not been cycled as often.

In the NOR flash world, wear leveling is typically implemented by an algorithm running on the host. This algorithm controls the mapping of logical data to physical blocks. It must change this mapping dynamically to ensure that wear leveling is effective.

For further reading, see the Wear Leveling Wikipedia article and its references.

5.4 Error Correction Codes

If the flash has the capability to execute error correction codes (ECCs), the endurance and data retention lifetime of the part can be extended. The ECC corrects errors in the data or code contained in the NOR Flash memory. The errors generated by endurance and data retention failures can be corrected by ECC.

Since ECC does not correct all possible errors, this feature will extend the lifetime of the memory, but not indefinitely. Once the errors become too numerous the likelihood of a multiple bit error in the corrected ECC unit increases and will eventually overcome the ability of the ECC to fix the errors.

ECC introduces area overhead due to the redundancy bits that are required for the error correction process.

There is a trade-off of chip area vs. lifetime which should be considered when designing a NOR Flash device.

6. Summary

This document has described the structure and operation of the NOR Flash transistor, its oxide degradation mechanisms, and the resultant effect on endurance and data retention. Test procedures and ideas for mitigating the effect of this degradation have been discussed. The level of detail has been restricted to avoid unnecessary complication of this subject.

7. References

A deeper understanding of these issues can be obtained through the following resources.

1. *JESD22-A117E- Electrically Erasable Programmable ROM (EEPROM) Program / Erase Endurance and Data Retention Stress Test*. JEDEC, 2024.
2. *JESD47L - Stress-Test-Driven Qualification of Integrated Circuits*. JEDEC, 2018.
3. *JEP122H - Failure Mechanisms and Models for Semiconductor Devices.* JEDEC, 2011.
4. S.M. Sze and Kwok K. Ng, *Physics of semiconductor Devices*. John Wiley and Sons, 2007.
5. E Cartier, “Characterization of the hot-electron-induced degradation in thin SiO_2 gate oxides.” In *Microelectronics Reliability,* Volume 38, Issue 2, 201-211. IBM Research Division, T. J. Watson Research Center, Yorktown Heights, 1998.
6. Young-Bog Park and D. K. Schroder, "Degradation of thin tunnel gate oxide under constant Fowler-Nordheim current stress for a flash EEPROM," in *IEEE Transactions on Electron Devices*, vol. 45, no. 6, pp. 1361-1368, June 1998.
7. AN500 NOR Flash Memory Erase Operation.
8. Wear Leveling Wikipedia article

8. Revision History

Revision	Date	Description
1.00	Mar 09, 2026	Initial release.

A. The Physics of Oxide Degradation

Hot electron injection and Fowler-Nordheim (FN) tunneling are responsible for much of the electron flow through the oxide. Most of the damage is caused by this conduction in the tunnel oxide layer and silicon-oxide interfaces during the program and erase cycles. The oxide physical damage is caused by repeated momentum transfer from the conduction electrons to the atoms in the tunnel oxide layer or at the silicon-oxide interfaces. Eventually more bonds will break between atoms, leaving traps that can hold charges.

An energy band diagram is an energy vs. distance illustration of electrons in a semiconductor device. The valence band is made of energy states of electrons bound to individual atoms, primarily silicon in the gate and substrate and silicon or oxygen in the tunneling oxide. Valence electrons do not move freely and so conduct very little electricity. The conduction band is formed by electrons shared by the silicon lattice and are free to move or conduct electricity. For the following figures, the left side is the floating gate (narrow band gap) of the NOR Flash device, the central portion (wide band gap) is the tunneling gate oxide, and the right side is the substrate (narrow band gap) or the gate channel. An electron can tunnel through this oxide under the proper conditions.

An energy band diagram under no bias between the floating gate and the substrate is described below. Under no bias the barrier is wide and the probability of an electron tunneling is low. Also, the probability of an electron tunneling in one direction is close to the probability of another electron tunneling in the opposite direction. So, you would expect little or no net current in either direction. Tunneling is a quantum mechanical phenomenon that allows electrons to pass through the oxide. If the energy barrier is high or wide, few electrons can tunnel through. If the barrier is low or narrow, there is a higher statistical probability that the electron will get through. Below it is a similar diagram after the bias is applied. If you apply bias, there is a higher probability of electrons tunneling in the favored direction than in the reverse direction. This is because electrons want to slide downhill in the conduction band just as if a ball were rolling downhill. It takes more energy to roll uphill, and so rolling downhill is a more likely outcome. You can see the bias creates a triangular region that allows electrons to tunnel through more easily because it is a narrower barrier. With applied bias you get easier tunneling and it becomes more likely electrons will tunnel in the favored direction since the barrier is higher in one direction than the other.

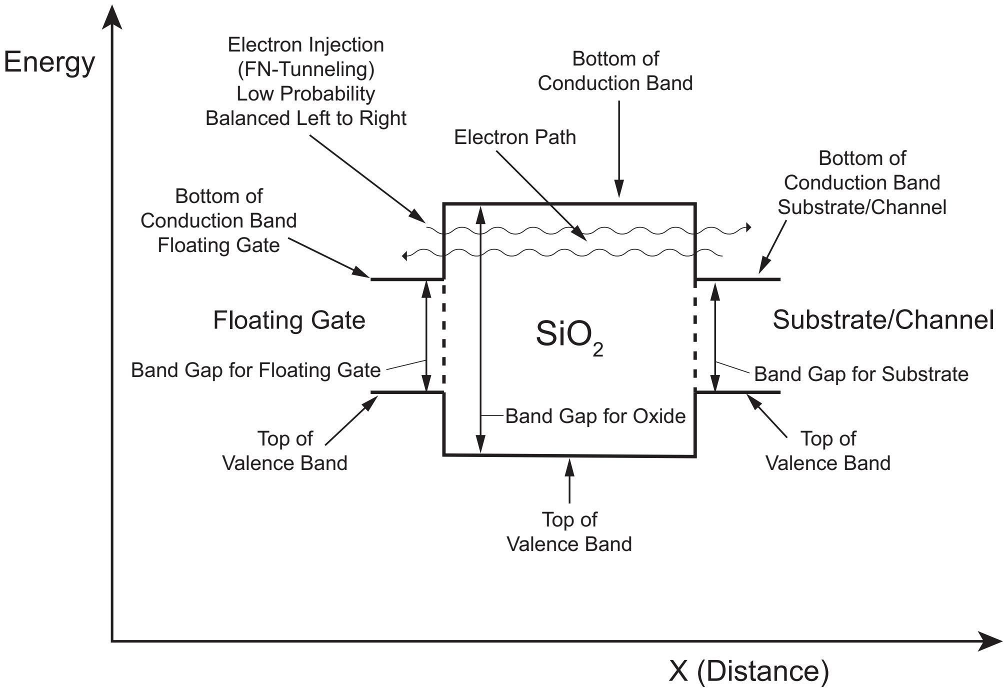


Figure 11. Unbiased Energy Band Diagram

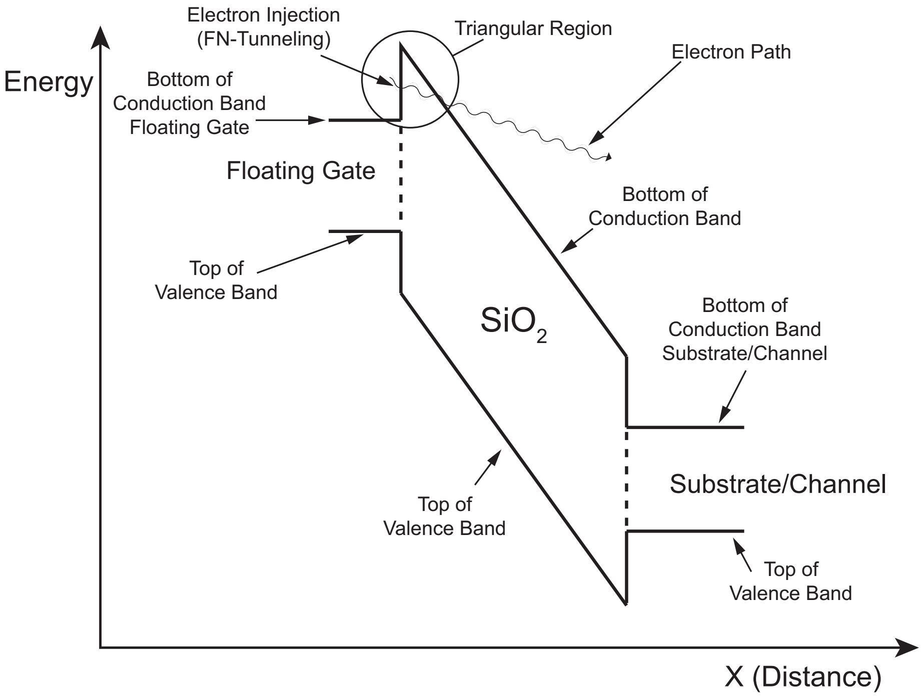


Figure 12. Voltage Biased Energy Band Diagram

Fowler-Nordheim tunneling is defined as electrons tunneling through the triangular region of the energy barrier caused by high-voltage potential between the floating gate and the substrate. This is shown in the energy band diagram in Figure 13.

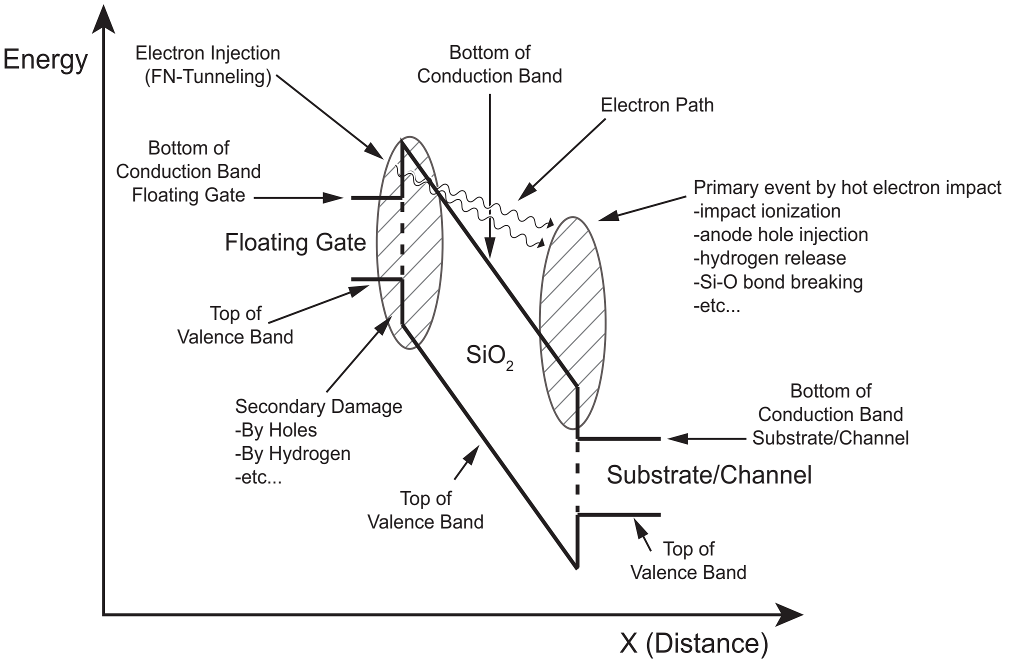


Figure 13. Homogenous stress using Fowler-Nordheim Injection

The tunneling electrons are accelerated in the oxide by the applied electric field. Hot-electron induced damage is located primarily near the substrate. This damage includes impact ionization in the oxide, hole damage, hydrogen release, possible Si-O bond breaking, etc. Damage will also occur near the floating gate even though fewer hot electrons are present there. This secondary damage is caused by hole trapping, electron-hole recombination and by chemical reactions of released impurities, such as hydrogen. The type of primary and secondary damage caused by this stress depends strongly on the electron energy near the substrate.[5]

Eventually leakage across the oxide increases as electrons no longer need to tunnel through the entire oxide. At that point electrons can tunnel the shorter distance between the trapping sites giving the electrons a higher probability of crossing the oxide for a given floating gate voltage. This leakage is known as Stress Induced Leakage Current or SILC. The increased tunneling between traps is called Trap Assisted Tunneling or TAT.

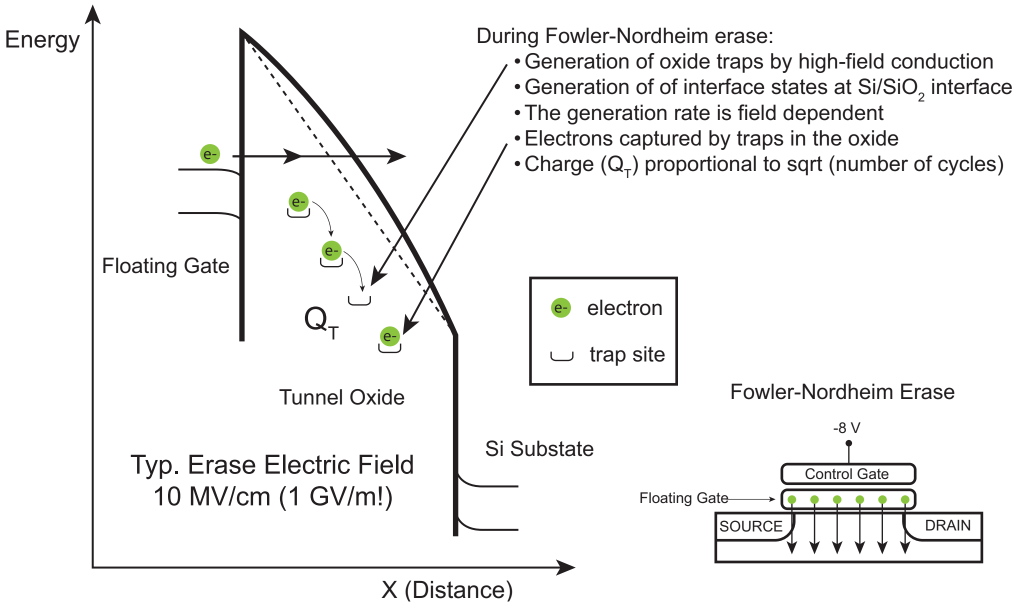


Figure 14. An Expanded View of the Tunnel Oxide with Electrons, Traps, and Electrons Tunneling Between Trap Sites

Endurance: Up to 100,000 Program Erase cycles for our devices.

This endurance number stated in our datasheets is an estimate of the number of program-erase cycles under a specific set of conditions; at a certain temperature range, and extended duty cycle, consistent with JEDEC standards. After the specified number of cycles, the likelihood that the transistor will change its characteristics and therefore cause the flash memory to fail the datasheet specification is increased. At some point this causes operational failure.

Note that the endurance number is not the number of cycles that the device can endure with constant use at maximum speed. The industry standard as directed by the JEDEC specification includes rest time for the traps to release some of the charges, and a more realistic use case similar to many customer applications.

As charges build up in the oxide, there is an effective screening of the floating gate voltage. Over time new states are created, and negative charges build up in the oxide. The charge screening raises the gate voltage needed to turn on the flash transistor (N-channel). This leads to a reduction in the transconductance or current output for a given gate voltage. Eventually the flash device will not perform the read/program/erase properly.

B. The Arrhenius Equation

The primary factors influencing the lifetime of a semiconductor device are described by the Arrhenius equation. This equation expresses the rate of chemical reaction as a function of temperature.

$$r = A\,e^{-\frac{E_a}{k}}$$

r = rate of the process

A = a proportional multiplier, which can be a function of temperature (*A* = *A(T)*)

Ea = a constant known as the activation energy for a given process. In practice this energy is an effective activation energy sometimes shown as *Eaa*.

k = Boltzmann's constant, 8.6 x 10^{-5} (eV/K)

T = absolute temperature in K

This equation has been adopted by the semiconductor industry to calculate lifetime of devices in varying temperature conditions. This allows semiconductor vendors to test devices at elevated temperatures for a relatively short time and use the equation to predict the device’s lifetime at normal operating temperature, typically a much longer period expressed in years.

By rearranging the Arrhenius equation, we can illustrate how the prediction is done. Assume a device passed a stress test for a period of t_{str} at an elevated temperature of T_{str}. If the expected normal usage temperature is T_{use}, we can calculate t_{use}, the equivalent time period the device is expected to function under these normal conditions.

$$t_{use} = t_{str}\,e^{-\frac{E_a}{k}\left(\frac{1}{T_{use}}-\frac{1}{T_{str}}\right)}$$

In this way, one stress test allows us to model a device’s lifetime under different conditions which the device may be used under in different applications.

IMPORTANT NOTICE AND DISCLAIMER

RENESAS ELECTRONICS CORPORATION AND ITS SUBSIDIARIES (“RENESAS”) PROVIDES TECHNICAL SPECIFICATIONS AND RELIABILITY DATA (INCLUDING DATASHEETS), DESIGN RESOURCES (INCLUDING REFERENCE DESIGNS), APPLICATION OR OTHER DESIGN ADVICE, WEB TOOLS, SAFETY INFORMATION, AND OTHER RESOURCES “AS IS” AND WITH ALL FAULTS, AND DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, OR NON-INFRINGEMENT OF THIRD-PARTY INTELLECTUAL PROPERTY RIGHTS.

These resources are intended for developers who are designing with Renesas products. You are solely responsible for (1) selecting the appropriate products for your application, (2) designing, validating, and testing your application, and (3) ensuring your application meets applicable standards, and any other safety, security, or other requirements. These resources are subject to change without notice. Renesas grants you permission to use these resources only to develop an application that uses Renesas products. Other reproduction or use of these resources is strictly prohibited. No license is granted to any other Renesas intellectual property or to any third-party intellectual property. Renesas disclaims responsibility for, and you will fully indemnify Renesas and its representatives against, any claims, damages, costs, losses, or liabilities arising from your use of these resources. Renesas' products are provided only subject to Renesas' Terms and Conditions of Sale or other applicable terms agreed to in writing. No use of any Renesas resources expands or otherwise alters any applicable warranties or warranty disclaimers for these products.

(Disclaimer Rev.1.01)

Corporate Headquarters

TOYOSU FORESIA, 3-2-24 Toyosu,
Koto-ku, Tokyo 135-0061, Japan
www.renesas.com

Contact Information

For further information on a product, technology, the most up-to-date version of a document, or your nearest sales office, please visit www.renesas.com/contact-us/.

Trademarks

Renesas and the Renesas logo are trademarks of Renesas Electronics Corporation. All trademarks and registered trademarks are the property of their respective owners.

© 2026 Renesas Electronics Corporation. All rights reserved.